

Distributed Compression and Squashed Entanglement

Ivan Savov

Master of Science

Physics Department

McGill University

Montreal, Quebec

February 6, 2008

A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of Master of Science.

©Ivan Savov, 2007

DEDICATION

To my parents, the only “system” I have respect for.

ACKNOWLEDGEMENTS

This work would not have been possible without the help and guidance of my supervisor Prof. Patrick Hayden. His outstanding pedagogical abilities and attention to detail have helped shape my understanding of the field of quantum information science at a world class level. In addition, I would like to thank Frédéric Dupuis, Sébastien Gambs and Omar Khalid for the many fruitful discussions about information theory and their help with the preparation of this manuscript. I also owe many thanks to Prof. David Avis and Leonid Chindelevitch for their assistance with some of the most difficult parts in this work. There are many other people who deserve an honorable mention for either directly or indirectly influencing me: Claude Crépeau, Aram Harrow, Debbie Leung, Jonathan Oppenheim and Andreas Winter. Last but not least, I want to thank my parents for cultivating in me the love of science and knowledge.

ABSTRACT

A single quantum state can be shared by many distant parties. In this thesis, we try to characterize the information contents of such distributed states by defining the multiparty information and the multiparty squashed entanglement, two steps toward a general theory of multiparty quantum information.

As a further step in that direction, we partially solve the multiparty distributed compression problem where multiple parties use quantum communication to faithfully transfer their shares of a state to a common receiver. We build a protocol for multiparty distributed compression based on the fully quantum Slepian-Wolf protocol and prove both inner and outer bounds on the achievable rate region. We relate our findings to previous results in information theory and discuss some possible applications.

ABRÉGÉ

Un état quantique peut être partagé entre plusieurs entités qui sont spatialement séparés. Dans ce mémoire, nous essayons de caractériser l'information quantique contenue dans de tels états distribués en définissant et utilisant les notions d'information multipartie (multiparty information) et d'intrication “écrasée” multipartie (multiparty squashed entanglement). Il s'agit de premiers pas vers une théorie générale de l'information quantique multipartie.

Nous faisons aussi un autre pas dans cette direction en étudiant le problème de la compression distribuée d'information quantique. En particulier, nous proposons un protocole de compression distribuée basé sur la version quantique du protocole de Slepian et Wolf et analysons ses caractéristiques. Nous discutons aussi la relation entre nos résultats et les travaux précédents dans la théorie de l'information et soulignons quelques applications possibles de notre protocole.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ABRÉGÉ	v
LIST OF FIGURES	viii
1 Introduction	1
2 Background	5
2.1 Classical information theory	5
2.1.1 Foundations	5
2.1.2 Shannon entropy	7
2.1.3 Typical sets	9
2.1.4 Compression	10
2.1.5 Multiple sources	12
2.1.6 Slepian-Wolf coding	13
2.2 Quantum information theory	16
2.2.1 Quantum states	16
2.2.2 von Neumann entropy	17
2.2.3 Quantum resources	20
2.2.4 Distance measures	22
2.2.5 Ensemble and entanglement fidelity	23
3 Results in quantum information theory	26
3.1 Schumacher compression	26
3.1.1 Typical subspace	27
3.1.2 Quantum compression	28
3.2 Quantum protocols as resource inequalities	29
3.2.1 The framework	29
3.2.2 The family of quantum protocols	30
3.3 State merging	32
3.3.1 Quantum conditional entropy	33
3.3.2 The state merging protocol	34
3.4 The fully quantum Slepian-Wolf protocol	37

	3.4.1	The protocol	38
	3.4.2	The FQSW resource inequality	40
4		Multiparty quantum information	43
	4.1	Multiparty information	46
	4.2	Multiparty squashed entanglement	49
	4.3	Example calculations of E_{sq}	56
5		Multiparty distributed compression	59
	5.1	The multiparty FQSW protocol	61
	5.1.1	Statement of results	62
	5.2	Proof of inner bound	64
	5.3	Proof of outer bound	75
	5.4	Discussion	82
6		Possible applications to the black hole information paradox	84
	6.1	Polygamy of purification	85
	6.2	Random internal dynamics for black holes	87
	6.3	Lost subsystem problem	88
7		Conclusion	90
		References	92

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
1–1 Dependency graph for the parts of this thesis.	4
2–1 Graphical representation of the conditional entropy and the mutual information.	13
2–2 The classical Slepian-Wolf rate region.	14
2–3 Quantum circuit illustrating the concept of entanglement fidelity.	24
3–1 Diagram of the state merging protocol.	33
3–2 Diagram of the ABR correlations before and after the FQSW protocol.	37
3–3 Circuit diagram for the FQSW protocol.	39
4–1 Entropy diagram for the multiparty information.	45
5–1 Representation of the quantum correlations in the multiparty distributed compression protocol.	60
5–2 The rate region for the multiparty FQSW protocol with three senders.	63
5–3 Two dimensional diagram showing the inner and outer bound on the rate region.	64
5–4 Detailed diagram of the distributed compression circuit.	76
6–1 Transfer of quantum correlations between three parties.	86
6–2 Black hole before and after emitting the radiation system R	87
6–3 Black hole which contains a lost subsystem L	89

CHAPTER 1

Introduction

Information theory is one of the most important mathematical theories developed in the last century. It finds applications in communications engineering, computer science, physics, economics, neuroscience and many other fields of modern science. Of particular interest are the recent developments in quantum information theory (QIT), a discipline which studies the limits that the laws of quantum mechanics impose on our ability to store, manipulate and transmit information. All information is physical; whether it be the magnetic domains of a hard disk platter, the reflective bumps on the surface of a DVD or the charge of the capacitors in a stick of RAM, that which we intuitively refer to as information must be stored in some physical system [1]. Thus, the incursion of quantum physics into information theory is inevitable if we want to understand the information properties of quantum systems like single photons and superconducting loops.

Modern quantum information theory has elaborated a paradigm in which a set of spatially localized parties try to accomplish a communication task by using communication resources like channels, states and quantum entanglement [2, 3, 4, 5]. Such an approach is now possible because of the substantial body of results characterizing quantum communication channels [6, 7, 8, 9] and the truly quantum resource of shared entanglement [10, 11, 12]. In this new quantum paradigm of information theory, many classical results need to be revisited in the light of the peculiar properties of quantum information.

In classical information theory, distributed compression is the search for the optimal rates at which two parties Alice and Bob can compress and transmit information faithfully to a third party Charlie. If the senders are allowed to communicate among themselves then they can obviously use the correlations between their sources to achieve better rates. The more interesting problem is to ask what rates can be achieved if no communication is allowed between the senders. The classical version of this problem was solved by Slepian and Wolf [13]. The quantum version of this problem was first approached in [14, 15] and more recently in [5], which describes the fully quantum Slepian-Wolf (FQSW) protocol and partially solves the distributed compression problem for two senders.

In this thesis, we analyze the multiparty scenario of distributed compression where many senders, Alice 1 through Alice m , send quantum information to a single receiver, Charlie. We will describe the multiparty FQSW protocol and exhibit a set of achievable rates for this protocol. We also derive an outer bound on the possible rates for *all* distributed compression protocols based on the multiparty squashed entanglement.

The multiparty squashed entanglement (independently discovered by Yang, et al. [16]) is a generalization of the squashed entanglement defined by Christandl and Winter [17] and has very desirable properties as a measure of multiparty entanglement. While there exist several measures for bipartite entanglement with useful properties and applications [2, 18, 19, 20], the theory of multiparty entanglement, despite considerable effort [21, 22, 23, 24], remains comparatively undeveloped. Multiparty entanglement is fundamentally more complicated because it cannot be described by a single number even for pure states. We can, however, define *useful* entanglement measures for particular

applications, and the multiparty squashed entanglement is one such measure well-suited to application in the distributed compression problem.

The main results of this thesis are contained in Chapters 4 and 5. Chapter 4 presents our original results on the multiparty generalization of squashed entanglement. Chapter 5 deals with the multiparty distributed compression problem and proves inner and outer bounds on the rate region. Before we get there, however, we will introduce some background material on classical and quantum information theory in Chapter 2. In Chapter 3, we describe some important recent results of quantum information theory which form the basic building blocks for our results. Finally, in Chapter 6 we take a look at some possible applications of the distributed compression results to the black hole information paradox. The dependency graph for the sections in this thesis is shown in Figure 1–1 on the next page.

Most of the original results in Chapters 4 and 5 appear in a paper [25] co-authored with Prof. David Avis and Prof. Patrick Hayden to which the author has made substantial contributions.

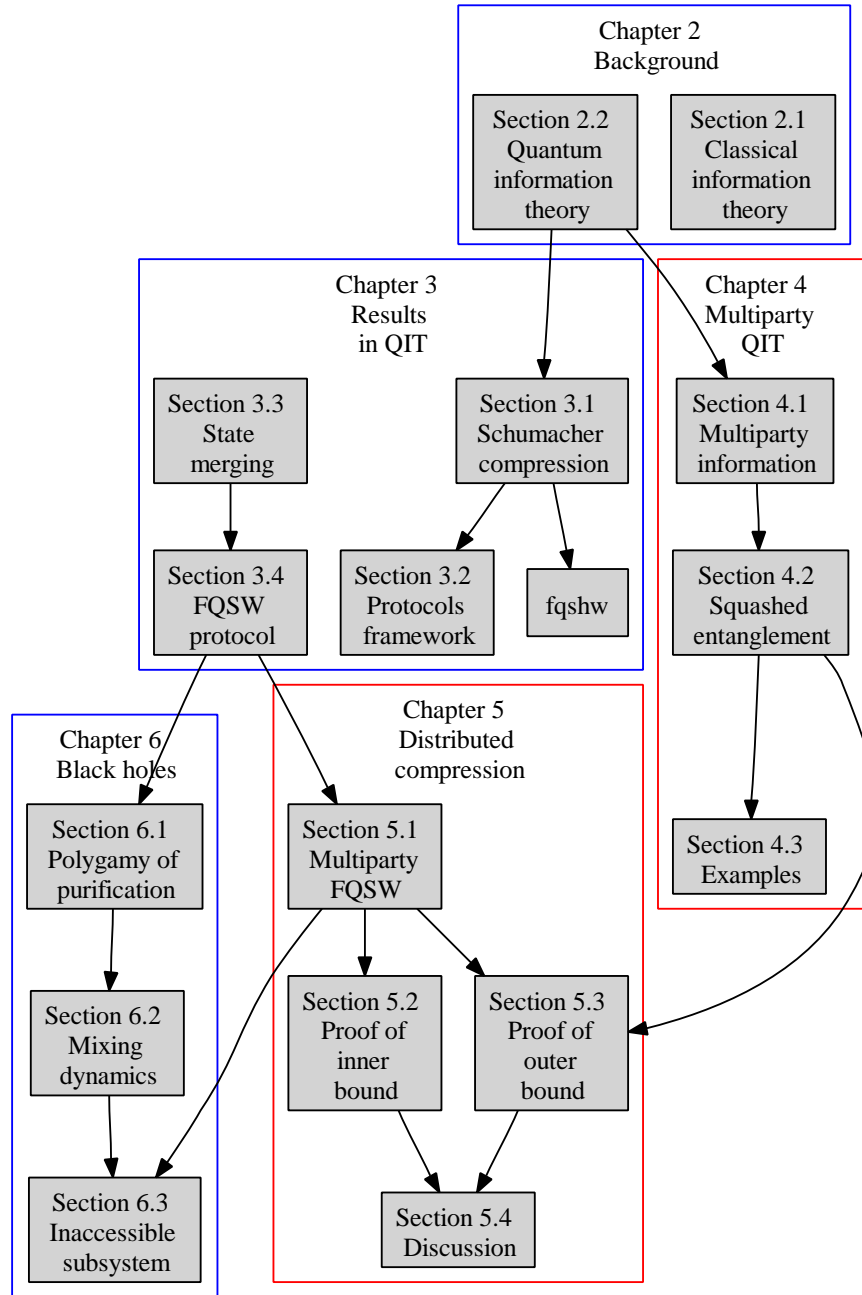


Figure 1–1: Dependency graph for the parts of this thesis.

CHAPTER 2

Background

In this chapter, we will present background concepts from classical information theory and their analogues in quantum information theory. These concepts form the basic building blocks with which we will construct all subsequent results. Our coverage of the information theoretic topics is far from exhaustive; it serves to introduce a minimum prerequisite structure that can support the rest of the exposition. For an in-depth view of classical and quantum information theory the reader is referred to the classics in the fields: [26] and [27] respectively.

2.1 Classical information theory

In 1948 Claude Shannon published a seminal paper [28] titled “A mathematical theory of communication” which set the stage for what has become one of the most fruitful modern mathematical theories. The field of *information theory* was born out of the need of communication engineers to quantify the information carrying capacities of channels and the theoretical aspects of data compression.

2.1.1 Foundations

At the root of Shannon’s information theory is the simplifying assumption that information ultimately boils down to the statistics of the symbols used to express it. Thus, another name for information theory could be *information statistics*. By focusing solely on the statistics of the symbols, which can be described by mathematical equations and axioms in the spirit of Hilbert’s

program [29], we can dispense with the difficult semantical questions related to humans.

We say that information is produced by a *source*, which is a random variable X that takes on values from an alphabet $\mathcal{X} = \{\alpha^1, \alpha^2, \dots, \alpha^{|\mathcal{X}|}\}$ according to some probability distribution $\Pr\{X = x\} = p(x)$.

Example 2.1. *Let X be the outcome of a coin flip. We will denote the alphabet $\mathcal{X} = \{‘H’, ‘T’\}$. If the coin is fair, then*

$$\Pr\{X = \alpha^1\} \equiv \Pr\{X = ‘H’\} = 0.5,$$

$$\Pr\{X = \alpha^2\} \equiv \Pr\{X = ‘T’\} = 0.5.$$

In this case, all outcomes are equally likely and it is maximally difficult to guess the result of the coin flip.

Example 2.2. *Suppose the Canadian border control center receives an hourly status message M from a distant outpost. The possible messages are:*

- *No one has attacked, which occurs 99.7% of the time: $\Pr\{M = \alpha^0\} = 0.997$*
- *The Americans have attacked: $\Pr\{M = \alpha^1\} = 0.002$*
- *The Russians have attacked: $\Pr\{M = \alpha^2\} = 0.001$ ¹*

In this scenario, one of the outcomes, α^0 , is much more likely than all the others. If we were to shut down the remote outpost and instead guess $M = \alpha^0$ every hour, we would only be wrong 0.3% of the time! Of course, implementing such an approximate border defense system is a silly idea, but in other situations an approximate result is just as good as the exact one.

¹ The quoted probabilities may not reflect the current geo-political balance of power.

Do we learn more information from one outpost message M , or from one coin flip X ? Using information theory, we should be able to *quantify* the amount of information produced by each source.

2.1.2 Shannon entropy

According to Shannon, there exists a single function sufficient to quantify the information content of a source. This function is the key building block in all of information theory.

Definition 2.3 (Shannon entropy). *Given a statistical source X , over the alphabet \mathcal{X} with probability function $p(x)$, the quantity*

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \quad (2.1)$$

is the Shannon entropy of the source.

The entropy of an unknown source measures our *uncertainty* about it and therefore, it measures how much information we learn, on average, when we look at a symbol from that source. The entropy is typically measured in *bits* since we use the base-2 logarithm in the calculation.

The quantity $-\sum_{x \in \mathcal{X}} p(x) \log p(x)$ also appears in thermodynamics where it is known as the Boltzmann-Gibbs entropy function. It is used to denote the logarithm of the number of available microstates that are consistent with certain macroscopic constraints [30, 31]. Together the entropy, energy, volume, pressure and temperature form the macroscopic description of a given thermodynamical system.

We now revisit the coin flip and outpost message scenarios from the previous examples.

Example 2.4. *The entropy of the balanced coin flip is:*

$$H(X) = -0.5 \log 0.5 - 0.5 \log 0.5 = -\log 0.5 = \log 2 = 1 \text{ [bit]}.$$

In other words, we learn one bit of information every time we flip the coin. On the other hand, the entropy of an outpost message is only

$$H(M) = -0.997 \log 0.997 - 0.002 \log 0.002 - 0.001 \log 0.001 = 0.03222 \text{ [bits]}.$$

Therefore, every coin flip carries about 30 times more information than a message from the distant outpost.

The true power of the information theoretic approach becomes apparent when we try to describe very long strings of symbols produced *independently* by the same source. Consider a source X which is used n times to produce the sequence X_1, X_2, \dots, X_n . We will denote the entire sequence as X^n with a superscript. We assume that the random variables X_i are independent and identically distributed (*i.i.d.*) according to $p(x)$.

We can write down the probability of a given string $x^n = x_1, x_2, \dots, x_n$ occurring as

$$\begin{aligned} \Pr\{X^n = x^n\} &= p(x_1, x_2, \dots, x_n) \\ &= p(x_1)p(x_2) \cdots p(x_n) \end{aligned} \tag{2.2}$$

since the X_i 's are independent.

Next we ask the important question:

“How often does the symbol α^i occur, on average, in a sequence of n uses of the source (X_1, \dots, X_n) ?”

Because every one of the symbols in the sequence has $\Pr\{X = \alpha^i\} = p(\alpha^i)$, the overall number of α^i 's in a string of length n is going to be approximatively

$np(\alpha^i)$. Therefore, on average, the probability of a string x_1, x_2, \dots, x_n is

$$p(x_1, \dots, x_n) = p(x_1)p(x_2) \cdots p(x_n) \quad (2.3)$$

$$\approx \underbrace{p(\alpha^1) \cdots p(\alpha^1)}_{np(\alpha^1) \text{ times}} \underbrace{p(\alpha^2) \cdots p(\alpha^2)}_{np(\alpha^2)} \cdots \underbrace{p(\alpha^{|\mathcal{X}|}) \cdots p(\alpha^{|\mathcal{X}|})}_{np(\alpha^{|\mathcal{X}|})} \quad (2.4)$$

$$\begin{aligned} &= \prod_{x \in \mathcal{X}} [p(x)]^{np(x)} \\ &= \prod_{x \in \mathcal{X}} \left[2^{\log_2 p(x)} \right]^{np(x)} \\ &= 2^{n[\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)]} \\ &= 2^{-nH(X)}. \end{aligned} \quad (2.5)$$

By going from equation (2.3) to (2.4), we have made a crucial change in our point of view: instead of taking into account the individual symbols x_i of the sequence, we focus on the global count of the symbol's occurrences. In other words, we abandon the microscopic description of the string and trade it for a macroscopic one in the spirit of thermodynamics. At first, it is difficult to believe that the typical sequences all have the same constant probability of occurrence, but we will see in the next section that this intuitive argument can be made rigorous.

2.1.3 Typical sets

Much of information theory is based on the concept of typical sequences. In the i.i.d. regime, nearly all of the sequences produced by the source have the same probability of occurrence. Consider the following theorem which makes precise our earlier argument.

Theorem 2.5 (Asymptotic equipartition theorem). *Let X_1, X_2, \dots, X_n be a sequence of independent random variables distributed according to $p(x)$, then*

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| -\frac{1}{n} \log p(X_1, X_2, \dots, X_n) - H(X) \right| > \epsilon \right\} = 0, \quad \forall \epsilon > 0. \quad (2.6)$$

In other words, for large enough n , the probability that a sequence occurs approaches $2^{-nH(X)}$ — a constant value. The result can also be interpreted in a different manner: sequences that have probability different from $2^{-nH(X)}$ are not likely to occur. Using this insight, we can partition the space of all possible sequences, \mathcal{X}^n , into two sets. The set of sequences that have probability of occurrence close to $2^{-nH(X)}$ and those that do not. We will call the former the set of typical sequences.

Definition 2.6 (Typical set). *The set of entropy typical sequences with respect to $p(x)$ is the set of all sequences x_1, x_2, \dots, x_n satisfying:*

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)} \quad (2.7)$$

we will denote this set $T_\epsilon^{(n)}$.

The typical set, $T_\epsilon^{(n)}$, has the following properties:

- (i) $\Pr\{X^n \in T_\epsilon^{(n)}\} \geq 1 - \delta \quad \forall \epsilon, \delta \text{ and } n \text{ sufficiently large.}$
- (ii) $|T_\epsilon^{(n)}| \leq 2^{n[H(X)+\epsilon]} \quad \forall \epsilon \text{ and } n \text{ sufficiently large.}$

Property (i) is a consequence of the asymptotic equipartition theorem and says that for large n , most of the sequences that come out of the source will be typical. Property (ii) is a bound on the size of the typical set which follows from the fact that all typical sequences occur with the same probability.

The bound on the size of the typical set is at the root of our ability to compress information.

2.1.4 Compression

Compression, also referred to as *source coding*, is our ability to encode a given source string into a shorter string while preserving most of the information contained therein. More generally, we talk about a *compression rate* which can be achieved for a given source X .

Definition 2.7 (Compression rate). *We say a compression rate R is achievable if for all $\epsilon > 0$, there exists $N(\epsilon)$ such that for $n > N(\epsilon)$, there exist maps:*

$$E_n : \mathcal{X}^n \rightarrow \mathcal{M} \quad |\mathcal{M}| = 2^{nR} \quad (2.8)$$

$$D_n : \mathcal{M} \rightarrow \mathcal{X}^n \quad (2.9)$$

such that

$$\Pr\{X^n \neq Y^n\} < \epsilon \quad (2.10)$$

where $Y^n = (D_n \circ E_n) X^n$.

Shannon's compression theorem [28] provides a bound the compression rates that are achievable for a given source X .

Theorem 2.8 (Shannon source coding). *Let $X^n \equiv X_1, X_2, \dots, X_n$ be a sequence of symbols i.i.d. $\sim p(x)$, then any compression rate R which satisfies*

$$R > H(X), \quad (2.11)$$

is achievable for n sufficiently large.

The idea behind Shannon compression is very simple. We begin by indexing the set of typical sequences $T_\epsilon^{(n)}$ in some order. We know that the size of $T_\epsilon^{(n)}$ is

$$|T_\epsilon^{(n)}| \leq 2^{n[H(X) + \epsilon]}. \quad (2.12)$$

therefore labels of length $\lceil H(X) + \epsilon \rceil$ bits will be sufficient to index the typical sequences. The encoding operation E_n for a given string x^n consists of:

- Recording the index of x^n if $x^n \in T_\epsilon^{(n)}$ and
- Rejecting the string and recording “error” if $x^n \notin T_\epsilon^{(n)}$.

The decoding operation D_n simply takes the index record and replaces it with the original string.

Because of Property (i) of the set of typical sequences, we know that the “error” condition will occur rarely:

$$\Pr\{x^n \notin T_\epsilon^{(n)}\} < \delta \quad \forall \epsilon, \delta \text{ and } n \text{ sufficiently large.} \quad (2.13)$$

This guarantees the low-error condition $\Pr\{X^n \neq Y^n\} < \delta$ for any $\delta > 0$. Shannon’s coding theorem holds since the rate $R = \lceil H(X) + \epsilon \rceil$ is achievable for any ϵ provided n is large enough.

2.1.5 Multiple sources

When we consider situations involving more than one source, some new information theoretic quantities become relevant. Consider now two sources X and Y distributed jointly according to $p(x, y)$. We will denote the marginals $p(x) = \sum_y p(x, y)$ and $p(y) = \sum_x p(x, y)$.

First we define the quantity

$$H(X|Y) = - \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(y)} \quad (2.14)$$

$$= \sum_y p(y) H(X|Y = y) \quad (2.15)$$

$$= H(XY) - H(Y) \quad (2.16)$$

which is known as *conditional entropy*. The conditional entropy measures the uncertainty in X that remains if we know the value of Y .

The quantity that quantifies how much information is shared between two sources is

$$I(X : Y) = H(X) + H(Y) - H(X, Y) \quad (2.17)$$

and is usually referred to as the *mutual information*. Two sources which have zero mutual information are independent.

The mutual information plays a key role in the characterization the information carrying capacity of memoryless channels. Together the compression

and channel capacity formulas are the two pillars of Shannon's information theory. In this thesis we focus mainly on compression problems and refer the reader interested in channel capacities to the classic texts [26, 32].

In order to get a better intuitive understanding of the conditional entropy and the mutual information, we often use a Venn-like diagram to represent them as in Figure 2-1.

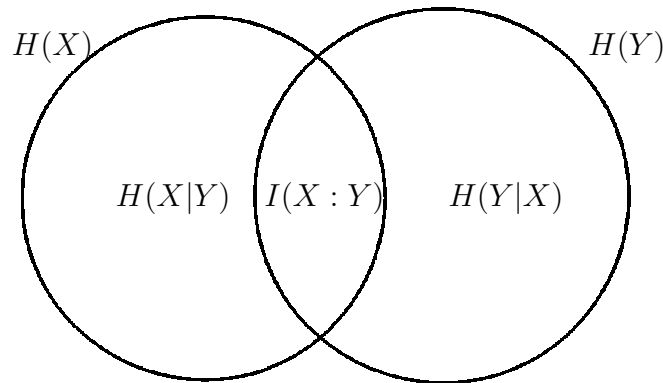


Figure 2-1: Graphical representation of the conditional entropy and the mutual information.

Furthermore, one can define the *conditional mutual information* by conditioning the mutual information formula on a third system Z .

$$I(X : Y|Z) = H(X|Z) + H(Y|Z) - H(XY|Z) \quad (2.18)$$

$$= H(XZ) + H(YZ) - H(XYZ) - H(Z). \quad (2.19)$$

The conditional mutual information measures the correlations between X and Y that are not shared with the variable Z .

2.1.6 Slepian-Wolf coding

Next we turn to the compression of two correlated sources X and Y distributed according to $p(x, y)$. If the two sources can be encoded together, then according to the Shannon's theorem a compression rate of $H(X, Y)$ can be achieved. The more interesting problem requires the sources to be encoded

separately without communication between the encoders. This is known as the Slepian-Wolf source coding problem [13].

Using the coding scheme suggested by Slepian and Wolf, we can compress at rates (R_X, R_Y) for X and Y respectively if they satisfy the inequalities

$$\begin{aligned} R_X &> H(X|Y), \\ R_Y &> H(Y|X), \\ R_X + R_Y &> H(XY). \end{aligned} \tag{2.20}$$

This set of inequalities corresponds to an achievable *rate region* in the (R_X, R_Y) -plane, as illustrated in Figure 2–2.

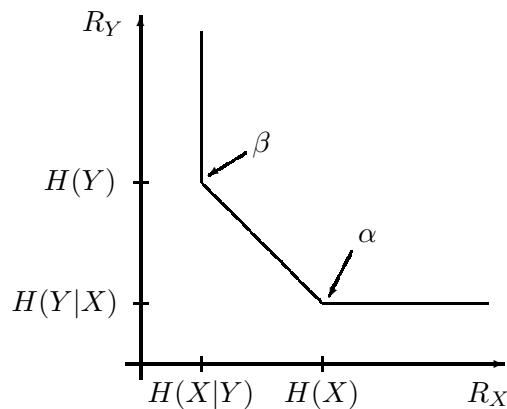


Figure 2–2: The classical Slepian-Wolf rate region. The points α and β are two corner points of the region.

To prove that the Slepian-Wolf rate region is achievable, we only need to show protocols which achieve the rates of the two corner points α and β . Any rate pair on the line between the two corner points can be achieved by *time sharing*. All other points in the rate region can be obtained by *resource wasting*.

The proof that the corner points are achievable relies on a coding scheme based on random bins and the properties of jointly typical sequences. A string

(x^n, y^n) is *jointly typical* if x^n is typical according to $p(x)$, y^n is typical according to $p(y)$ and (x^n, y^n) is typical according to $p(x, y)$. To encode, we will randomly assign to each string x^n an index $i(x^n) \in \{1, 2, \dots, 2^{nR_X}\}$. Similarly, to each y^n we assign an index $j(y^n) \in \{1, 2, \dots, 2^{nR_Y}\}$. The decoding operation takes the received indices (i, j) and tries to reproduce a copy of the original string (\hat{x}^n, \hat{y}^n) .

In the case of point α from Figure 2–2, the rates correspond to

$$R_X = H(X) + \epsilon_1, \quad (2.21)$$

$$R_Y = H(Y|X) + \epsilon_2. \quad (2.22)$$

in the limit where ϵ_1 and ϵ_2 go to zero. According to Shannon's source coding theorem (Theorem 2.8), the rate of equation (2.21) is sufficient to faithfully decode the string x^n , i.e. with high probability $\hat{x}^n = x^n$. The decoder then has to find the string y^n which is jointly typical with the decoded \hat{x}^n and this is possible provided the rate R_Y is greater than the conditional entropy $H(Y|X)$. The coding scheme for point β is analogous.

The multiparty version of the Slepian-Wolf problem was considered in [33, 34]. In the multiparty case, we have not two but m sources X_1, X_2, \dots, X_m which are to be encoded separately and decoded by a common receiver. We want to know the optimal rate tuple (R_1, R_2, \dots, R_m) at which we can compress the corresponding sources such that the information can be recovered faithfully after decoding. It is shown in [34] that the rates have to satisfy the following set of inequalities

$$\sum_{k \in \mathcal{K}} R_k > H(X_{\mathcal{K}} | X_{\bar{\mathcal{K}}}), \quad (2.23)$$

for all $\mathcal{K} \subseteq \{1, 2, \dots, m\}$, $\bar{\mathcal{K}} = \{1, 2, \dots, m\} \setminus \mathcal{K}$ and $X_{\mathcal{K}} := \{X_i : i \in \mathcal{K}\}$. Note that the two-party inequalities (2.20) are a special case of the more general multiparty result.

2.2 Quantum information theory

The fundamental ideas of quantum information theory are analogous to those of classical information theory. In addition to the classical sources and channels, we simply introduce a new set of fundamental building blocks in our studies. These *quantum resources* governed by the laws of quantum mechanics can exhibit strange and non-intuitive behaviour but can nevertheless be studied with the techniques of information theory.

2.2.1 Quantum states

The fundamental principles of quantum mechanics are simple enough to be explained in the space available on the back of an envelope, but to truly understand the implications of these principles takes years of training and effort. We assume the reader is familiar with basic notions of quantum mechanics [35, 27]. This section will focus on specific notions and notation that are used in quantum information theory.

We will denote quantum systems by uppercase roman letters like A, B, R and the corresponding Hilbert spaces as $\mathcal{H}^A, \mathcal{H}^B, \mathcal{H}^R$ with respective dimensions d_A, d_B, d_R . We denote pure states of the system A by *kets*: $|\varphi\rangle^A$ and *density matrices* as φ^A . Because of the probabilistic interpretation of quantum mechanics, all kets have unit norm and all density matrices are positive and with unit trace. We will refer to both kets and density matrices as *states*.

We use the partial trace operator to model partial knowledge of a state. Given a bipartite state ρ^{AB} shared between Alice and Bob, we say that Alice holds in her lab the reduced density matrix: $\rho^A = \text{Tr}_B \rho^{AB}$, where Tr_B denotes

a partial trace over Bob's degrees of freedom. In general the state produced in this manner will be *mixed* – a classical probability distribution over states.

Conversely, any mixed state $\sigma^A \in \mathcal{H}^A$ can be *purified* to a fictitious larger Hilbert space. That is, we imagine a corresponding pure state $|\sigma\rangle^{AR} \in \mathcal{H}^A \otimes \mathcal{H}^R$ such that taking the partial trace over the R system gives the original state: $\text{Tr}_R(|\sigma\rangle\langle\sigma|^{AR}) = \sigma^A$. The purification procedure is often referred to as escaping to the *Church of the larger Hilbert space* in literature.

2.2.2 von Neumann entropy

Analogously to classical information theory, we quantify the information content of quantum systems by using an entropy function.

Definition 2.9 (von Neumann Entropy). *Given the density matrix $\rho^A \in \mathcal{H}^A$, the expression*

$$H(A)_\rho = -\text{Tr}(\rho^A \log \rho^A) \quad (2.24)$$

is known as the von Neumann entropy of the state ρ^A .

Certain texts use the alternate notation $S(A)_\rho$ for the von Neumann entropy to distinguish it from the classical Shannon entropy, but we choose not to make this distinction here. This overloading of notation is warranted since the von Neumann entropy is in fact the Shannon entropy of the eigenvalues of the state. Given the spectral decomposition of the state $\rho^A = \sum_i \lambda_i |e_i\rangle\langle e_i|$, we can calculate $H(A)_\rho = -\text{Tr}(\rho^A \log \rho^A) = -\sum_i \lambda_i \log \lambda_i$. The von Neumann entropy of a pure state is zero, since it has only a single eigenvalue.

² Strictly speaking, we should say $\sigma^A \in D(\mathcal{H}^A)$ where $D(\mathcal{H}^A)$ is the set of density matrices over \mathcal{H}^A . We will use this economy of notation consistently.

For bipartite states ρ^{AB} we can also define the quantum conditional entropy

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho \quad (2.25)$$

where $H(B)_\rho = -\text{Tr}(\rho^B \log \rho^B)$ is the entropy of the reduced density matrix $\rho^B = \text{Tr}_A(\rho^{AB})$. In the same fashion we can also define the quantum mutual information

$$I(A; B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho \quad (2.26)$$

and in the case of a tripartite system ρ^{ABC} we define the conditional mutual information as

$$I(A; B|C)_\rho := H(A|C)_\rho + H(B|C)_\rho - H(AB|C)_\rho \quad (2.27)$$

$$= H(AC)_\rho + H(BC)_\rho - H(ABC)_\rho - H(C)_\rho. \quad (2.28)$$

It can be shown that $I(A; B|C)$ is strictly non negative for any state ρ^{ABC} . The formula $I(A; B|C) \geq 0$ can also be written in the form

$$H(AC) + H(BC) \geq H(C) + H(ABC). \quad (2.29)$$

This inequality, originally proved in [36], is called the *strong subadditivity* of von Neumann entropy and forms an important building block of quantum information theory.

On the surface, it may appear to the reader that quantum information theory has nothing new to offer except a rewriting of the classical formulas in a new context. This observation is highly misleading. We present the following example to illustrate some of the new aspects of quantum information theory.

Example 2.10. *Consider the Φ^+ Bell state*

$$|\Phi\rangle^{AB} = \frac{1}{\sqrt{2}}(|00\rangle^{AB} + |11\rangle^{AB}). \quad (2.30)$$

This state exhibits a form of quantum correlation called entanglement that is fundamentally different from classical correlation. The associated density matrix is $\Phi^{AB} = |\Phi\rangle\langle\Phi|^{AB}$, which has the reduced density matrices $\Phi^A = \Phi^B = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$.

Next we calculate the entropy of the two subsystems A , B and the system as a whole

$$H(A)_\Phi = 1, \quad H(B)_\Phi = 1, \quad H(AB)_\Phi = 0, \quad (2.31)$$

since Φ^A, Φ^B are maximally mixed and $|\Phi\rangle^{AB}$ is pure. Using these results, it is now simple to calculate the conditional entropy

$$H(A|B) = H(AB) - H(B) = -1 \text{ [bits]}, \quad (2.32)$$

and the mutual information

$$I(A; B) = H(A) + H(B) - H(AB) = 2 \text{ [bits]}. \quad (2.33)$$

Equation (2.32) illustrates one of the key differences between classical information theory and quantum information theory: the fact that conditional entropy can be negative. How can we interpret negative values as uncertainties? Also, it is not immediately clear what we mean by conditioning on a quantum system in the first place. These issues will be discussed in some detail in Section 3.3 where we will give the conditional entropy an operational interpretation.

In classical information theory, the mutual information between two binary sources attains its maximal value of 1 when the two sources are perfectly correlated. As we can see from equation (2.33), in the quantum world two qubits can be, in some sense, *more than perfectly correlated* and have mutual information as much as 2 bits!

2.2.3 Quantum resources

The current trend in quantum information theory is to look at communication tasks as inter-conversions between clearly defined information resources. To render the resource picture generic, we always imagine a scenario in which two localized parties, usually called Alice and Bob, want to perform a certain communication task. Local computation will be regarded as free of cost in order to focus on the communication aspects of the task.

An example of a classical communication resource is the *noiseless channel* from Alice to Bob, denoted $[c \rightarrow c]$. The symbol $[c \rightarrow c]$ represents the ability to send one bit of information from Alice to Bob. A related classical resources is the *noisy channel*, denoted $\{c \rightarrow c\}$ which is usually modeled as a mapping $\mathcal{N}^{X \rightarrow Y}$, described by a conditional probability $p(Y = y|X = x)$ where X is the input variable sent by Alice and Y the random variable received by Bob. The noiseless channel $[c \rightarrow c]$ is, therefore, a special case of the general channel $\{c \rightarrow c\}$ with the identity mapping $\mathcal{N} = \mathbf{1}^{X \rightarrow Y}$ from X to Y . Another classical resource denoted $[cc]$ represents a random bit shared between Alice and Bob.

Quantum information theory introduces a new set of resources. In analogy to the classical case, we have the *noiseless quantum channel* $[q \rightarrow q]$ which represents the ability to transfers one *qubit*, a generic two dimensional quantum system, from Alice to Bob. A *noisy quantum channel*, $\{q \rightarrow q\}$, is modeled by a mapping $\mathcal{N}^{A \rightarrow B}$ which takes density matrices in \mathcal{H}^A to density matrices in \mathcal{H}^B . The mapping \mathcal{N} is a *quantum operation*: a completely positive trace preserving (CPTP) operator [27].

One key new resource of quantum information theory is the maximally entangled state shared between Alice and Bob

$$|\Phi\rangle^{AB} = \frac{1}{\sqrt{2}}(|00\rangle^{AB} + |11\rangle^{AB}), \quad (2.34)$$

which we denote $[qq]$. Note that, since local operations are allowed for free in our formalism, any state $|\Phi'\rangle^{AB} = U^A \otimes U^B |\Phi\rangle^{AB}$ where U^A, U^B are local unitary operations is equivalent to $|\Phi\rangle^{AB}$. Entanglement is a fundamental quantum resource because it cannot be generated by local operations and classical communication (LOCC). The precise characterization of entanglement has been a great focal point of research in the last decade. For an in depth review of the subject we refer the readers to the excellent papers [20, 37].

Entanglement forms a crucial building block for quantum information theory because it can be used to perform or assist with many communication tasks. In particular, two of the first quantum protocols that ever appeared involve *ebits*, or entangled bits. The *quantum teleportation* protocol [38] uses entanglement and two bits of classical communication to send a quantum state from Alice to Bob

$$[qq] + 2[c \rightarrow c] \geq [q \rightarrow q], \quad (\text{TP})$$

while the *superdense coding* protocol [39] uses entanglement to send two classical bits of information with only a single use of a quantum channel

$$[qq] + [q \rightarrow q] \geq 2[c \rightarrow c]. \quad (\text{SC})$$

The above *resource inequalities* indicate that the resources on the left hand side can be used to simulate the resource on the right hand side.

The two protocols (TP) and (SC) are only the tip of the iceberg: there are many more protocols and fundamental results in quantum information theory that can be written as resource inequalities. In Section 3.2 we will introduce some of them and the relationships that exist between them.

2.2.4 Distance measures

In order to describe the “distance” between two quantum states we use the notions of *trace distance* and *fidelity*. The trace distance between quantum states σ and ρ is

$$TD(\rho, \sigma) := \|\rho - \sigma\|_1 = \text{Tr}|\rho - \sigma| \quad (2.35)$$

where $|X| = \sqrt{X^\dagger X}$.

The fidelity between two pure states is simply the square of their inner product

$$F(|\varphi\rangle, |\psi\rangle) = |\langle\varphi|\psi\rangle|^2. \quad (2.36)$$

The most natural generalization of this notion to mixed states ρ, σ is the formula

$$F(\rho, \sigma) = \text{Tr} \left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2. \quad (2.37)$$

Two states that are very similar have fidelity close to 1 whereas states with little similarity will have low fidelity.

Note that some texts, (ex: [27]) define the trace distance with an extra normalization factor of $\frac{1}{2}$ and write the fidelity without the square. These differences of convention do not affect any of our findings but are important to point out to avoid confusion.

The trace distance and fidelity measures are related, that is if two states ρ and σ are close in one measure they are also close in the other [40]. More precisely, the quantities TD and F satisfy the following inequalities

$$1 - \sqrt{F} \leq \frac{1}{2}TD \leq \sqrt{1 - F}, \quad (2.38)$$

$$1 - TD \leq F \leq 1 - \frac{TD^2}{4}. \quad (2.39)$$

Thus, if for certain states $F \geq 1 - \epsilon$, then $TD \leq 2\sqrt{\epsilon}$. Also, if $TD \leq \epsilon$, then $F \geq 1 - \epsilon$.

2.2.5 Ensemble and entanglement fidelity

The concept of an identical, independently distributed (i.i.d.) source also exists in quantum information theory. However, there are a number of ways we can adapt the concept to the quantum setting so some clarifications are in order.

An ensemble $\mathcal{E} = \{p_i, |\psi_i\rangle\}$ is a set of quantum states $|\psi_i\rangle$ which occur with probability p_i . One way to describe a quantum source is to specify the states $|\psi_i\rangle$ and the corresponding probabilities p_i associated with this source. Using this ensemble characterization we can specify what it means to successfully perform a communication protocol with that source. Let $\mathcal{N}^{A \rightarrow \hat{A}}$ with input $|\psi\rangle^A \in \mathcal{H}^A$ and output $\sigma^{\hat{A}} \in \mathcal{H}^{\hat{A}}$ be the quantum operation associated with the protocol:

$$\mathcal{N}(|\psi\rangle\langle\psi|) = \sigma^{\hat{A}}. \quad (2.40)$$

To measure how faithfully the input state has been reproduced at the output we calculate the input-output fidelity $F(|\psi\rangle^A, \sigma^{\hat{A}})$. In order to measure how faithfully the source as a whole is reproduced at the output, we have to average over the input-output fidelities of the ensemble

$$\bar{F}(\mathcal{E}, \mathcal{N}) := \sum_i p_i F(|\psi_i\rangle, \sigma_i), \quad \sigma_i = \mathcal{N}(|\psi_i\rangle\langle\psi_i|). \quad (2.41)$$

If we want the source to be preserved perfectly then we require $\bar{F}(\mathcal{E}, \mathcal{N}) = 1$. In general, however, we will be content with approximate transmission where

$$\bar{F}(\mathcal{E}, \mathcal{N}) \geq 1 - \epsilon \quad (2.42)$$

for arbitrary small ϵ . It turns out that this way of describing the source may not be practical or desirable since it requires a detailed knowledge of the inner workings of the source — something that is often impossible to obtain even in theory.

The better way to describe a quantum source is specify only the average density operator $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ for that source. This characterization could be obtained through *state tomography* [27] and does not presuppose any knowledge of the ensemble which generates ρ . This description is more general because the results we obtain for the density matrix ρ will hold for *all* ensembles $\{p_i, |\psi_i\rangle\}$ such that $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$.

This also leads us to an alternative and simpler way of judging the success of a quantum protocol that relies on the idea of the *Church of the larger Hilbert space*. Let $|\psi\rangle^{AR}$ be a purification of ρ^A to some reference system R . This reference system is entirely fiducial and does not participate in the protocol. In the larger Hilbert space $\mathcal{H}^A \otimes \mathcal{H}^R$ the $\mathcal{N}^{A \rightarrow \hat{A}}$ operation acts as

$$\mathcal{N}^{A \rightarrow \hat{A}} \otimes \mathbf{1}^R (|\psi\rangle\langle\psi|^{AR}) = \sigma^{\hat{A}R}. \quad (2.43)$$

The operation is shown as a quantum circuit in Figure 2–3.

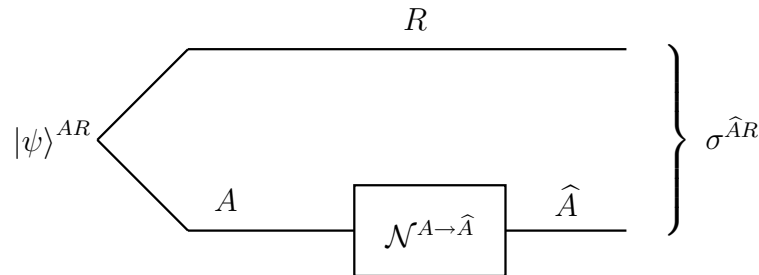


Figure 2–3: A quantum circuit which shows \mathcal{N} acting on the A system while the reference, R , is left unperturbed.

For approximate transmission, we now require the fidelity between the pure input state $|\psi\rangle^{AR}$ and the possibly mixed output state $\sigma^{\hat{A}R}$ to be high

$$F(|\psi\rangle^{AR}, \sigma^{\hat{A}R}) = \langle\psi^{AR} | \sigma^{\hat{A}R} | \psi^{AR}\rangle \geq 1 - \epsilon. \quad (2.44)$$

Equation (2.44) measures the *entanglement fidelity* of the operation: how well the protocol manages to transfers the R -entanglement from the A system to

the \hat{A} system. It can be shown [41] that if the channel \mathcal{N} has high entanglement fidelity then the average fidelity $\bar{F}(\mathcal{E}, \mathcal{N})$ will also be high for any ensemble \mathcal{E} such that $\rho^A = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. In other words, equation (2.44) implies equation (2.42). The entanglement fidelity paradigm has the advantage that the input state to the protocol is pure, which makes our analysis much simpler. Also, in this paradigm we are certain that any correlations the A system might have with other systems are preserved because of monogamy of entanglement.

In the i.i.d. setting, we operate simultaneously on n copies of the same input state ρ^A . We denote the tensor product of all the input states by $\rho^{A^n} = \rho^A \otimes \cdots \otimes \rho^A$ (n -copies). The quantum operation becomes $\mathcal{N}^{A^n \rightarrow \hat{A}^n}$ and the output state will be $\sigma^{\hat{A}^n}$. The entanglement fidelity

$$F(|\psi\rangle^{A^n R^n}, \sigma^{\hat{A}^n R^n}) = \langle \psi^{A^n R^n} | \sigma^{\hat{A}^n R^n} | \psi^{A^n R^n} \rangle \geq 1 - \epsilon(n), \quad (2.45)$$

is now a function of n , the *block size* of the protocol. Thus, in the i.i.d. setting we say the protocol implemented by \mathcal{N} succeeds when $\epsilon(n) \rightarrow 0$ as $n \rightarrow \infty$. More formally, for any required precision ϵ_0 , there exists an $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, there exist n -dependent maps \mathcal{N} such that $\epsilon(n) < \epsilon_0$ in equation (2.45).

CHAPTER 3

Results in quantum information theory

This chapter is dedicated to four landmark results in quantum information theory. The first of these is Schumacher compression, the quantum version of source coding [42]. The second is the *resource framework of quantum information theory* [4], which defines rigorously the properties of quantum protocols and discusses the relationships between them [3]. Then, in section 3.3, we focus our attention on one protocol for compression of quantum information with side information known as *state merging* [15, 43]. Finally, in the last section of this chapter, we discuss in detail the fully quantum Slepian-Wolf (FQSW) protocol for state transfer and simultaneous entanglement distillation. To a large extent, the multiparty results in this thesis are a direct generalization of the two-party FQSW protocol, therefore, section 3.4 is of central importance to the remainder of the argument.

3.1 Schumacher compression

If classical information theory is 60 years old [28] then quantum information theory must be 12 years old. Indeed, we can say that Schumacher laid the foundations of quantum information theory with his 1995 paper [42] where he showed that the von Neumann entropy, $H(\rho)$, plays the analogous role of Shannon entropy for quantum systems. Namely, it has operational interpretation as the number of qubits necessary to convey the information from a quantum source ρ .

3.1.1 Typical subspace

The notion of a typical set (section 2.1.3) can easily be generalized to the quantum setting. Consider a source which produces many copies of the state ρ^A which has spectral decomposition $\rho^A = \sum_i \lambda_i |i\rangle\langle i|$.

In the i.i.d. regime, the state produced by the source is given by $\rho^{A^n} = \rho^{A_1} \otimes \rho^{A_2} \otimes \dots \otimes \rho^{A_n}$ which can be written as

$$\begin{aligned} \rho^{A^n} &= \sum_{i^n} \lambda_{i_1} \dots \lambda_{i_n} |i_1\rangle\langle i_1|^{A_1} \otimes \dots \otimes |i_n\rangle\langle i_n|^{A_n} \\ &= \sum_{i^n} \lambda_{i^n} |i^n\rangle\langle i^n|^{A^n}. \end{aligned} \tag{3.1}$$

We now define the *typical projector* as follows

$$\begin{aligned} \Pi_\epsilon^{(n)} &= \sum_{i^n \in T_\epsilon^{(n)}} |i_1\rangle\langle i_1|^{A_1} \otimes |i_2\rangle\langle i_2|^{A_2} \otimes \dots \otimes |i_n\rangle\langle i_n|^{A_n} \\ &= \sum_{i^n \in T_\epsilon^{(n)}} |i^n\rangle\langle i^n|^{A^n}, \end{aligned} \tag{3.2}$$

where we sum over all the typical sequences $T_\epsilon^{(n)}$ with respect to the classical probability distribution $p(i) := \lambda_i$.

We call the support of $\Pi_\epsilon^{(n)}$, the *typical subspace* of \mathcal{H}^{A^n} associated with ρ^A . The typical subspace, by its construction, inherits the characteristics of the typical set. Indeed, $\Pi_\epsilon^{(n)}$ has the following properties

- (i) $\text{Tr} \left[\rho^{A^n} \Pi_\epsilon^{(n)} \right] > 1 - \delta \quad \forall \delta, \epsilon > 0 \text{ and } n \text{ sufficiently large.}$
- (ii) $\text{Tr} \left[\Pi_\epsilon^{(n)} \right] \leq 2^{n[H(A)_\rho + \epsilon]} \quad \forall \epsilon > 0 \text{ and } n \text{ sufficiently large.}$

Property (i) says that, for large n , most of the states produced by the source will lie mostly inside the typical subspace. Property (ii) is a bound on the size of the typical subspace which follows from the classical bound on the size of the typical set $T_\epsilon^{(n)}$. These two properties are at the heart of our ability to compress quantum information.

3.1.2 Quantum compression

Analogously to the classical case, we have the notion of a quantum compression rate. In the quantum regime, we use the entanglement fidelity (see Section 2.2.5) to measure how well the state is reproduced after decoding.

Definition 3.1 (Quantum compression rate). *We say a compression rate R for the source ρ^A is achievable if for all ϵ , there exists $N(\epsilon)$ such that for $n > N(\epsilon)$, there exist maps:*

$$\mathcal{E}_n : \mathcal{H}^{A^n} \rightarrow \mathcal{M} \quad |\mathcal{M}| = 2^{nR} \quad (3.3)$$

$$\mathcal{D}_n : \mathcal{M} \rightarrow \mathcal{H}^{\hat{A}^n} \quad (3.4)$$

such that the purification $|\psi\rangle^{A^n R^n}$ of ρ^{A^n} satisfies

$$F(|\psi\rangle^{A^n R^n}, \sigma^{\hat{A}^n R^n}) = \langle \psi | \sigma^{\hat{A}^n R^n} | \psi \rangle^{A^n R^n} > 1 - \epsilon. \quad (3.5)$$

where $\sigma^{\hat{A}^n R^n} = \mathcal{D}_n \circ \mathcal{E}_n \otimes \mathbf{1}^{R^n} (|\psi\rangle\langle\psi|)$.

Theorem 3.2 (Schumacher noiseless coding). *An i.i.d. quantum source ρ^A can be compressed at a rate R if $R > H(A)_\rho$ and cannot if $R < H(A)_\rho$.*

The idea behind the Schumacher compression result is simple. We encode by performing the measurement

$$M_{\mathcal{E}} = \{\Pi_{\epsilon}^{(n)}, \mathbf{1} - \Pi_{\epsilon}^{(n)}\}. \quad (3.6)$$

If $\Pi_{\epsilon}^{(n)}$ occurs, we keep this state since it is typical. Otherwise, if $(\mathbf{1} - \Pi_{\epsilon}^{(n)})$ occurs, we replace the state with some fixed state $|\text{err}\rangle$ as an indicator that an error has occurred. The decoding operation \mathcal{D}_n is the identity operation. Property (i) from the previous section guarantees that the probability of error tends to zero when n becomes large. Also, since we only send states within the typical subspace, Property (ii) gives us a bound on the amount of quantum information necessary to convey this state.

Note that the compression protocol described above works both for scenarios where the mixed state is obtained from a stochastic average over pure states $\rho^A = \sum_i p_i |\psi_i\rangle\langle\psi_i|^A$ and scenarios where the density matrix is part of a larger pure state $\rho^A = \text{Tr}_E |\Psi\rangle\langle\Psi|^{AE}$.

3.2 Quantum protocols as resource inequalities

Most of the old results of quantum information theory form a loose collection of coding theorems, each of them with applications only to one specific communication task. Recently, there has been a push to organize these results into a unified framework of resource inequalities [2, 3, 4, 5]. A resource inequality is a quantitative statement regarding inter-conversions between clearly defined *generic* information resources. The key benefit of such a framework is that, like LEGO blocks, we can build one communication protocol based on another, and generally work at a higher level of abstraction than is possible when working with the specifics of each protocol.

3.2.1 The framework

A unified framework for both classical and quantum information theory was developed in [4]. The notions “resource” and “protocol” are clearly defined as well as the rules for combining and composing them. In particular, this framework deals with the class of bipartite, unidirectional communication tasks involving memoryless channels and sources in the i.i.d regime.

Borrowing from the cryptography heritage, the two main participants in the protocols are called Alice and Bob. Alice is usually the sender, and performs some encoding operation while Bob does the decoding. Additionally, the framework introduces two novel participants *Eve* and the *Reference*. We use Eve to model information lost to the environment in a noisy channel. The reference R is a fiducial purification system which allows us to deal with mixed

states in a simple manner as discussed in Section 2.2.5. Most important of all, the framework introduces the *Source*, which produces some state ρ^S and distributes it to the participants before the beginning of the protocol.

In Section 2.2.3, we introduced some of the resources of information theory like the noiseless classical channel $[c \rightarrow c]$, noisy classical channel $\{c \rightarrow c\}$ and the quantum equivalents $[q \rightarrow q]$ and $\{q \rightarrow q\}$. In order to be more precise, we sometimes use a different notation for noisy channels

$$\{q \rightarrow q\} \equiv \langle \mathcal{N} \rangle \quad (3.7)$$

which explicitly shows the map \mathcal{N} associated with that channel. Note that the angle brackets $\langle . \rangle$ indicate that we are working in the asymptotic regime of many copies of the resource: $\langle \mathcal{N} \rangle \sim \mathcal{N}^{\otimes n}$ and $\langle \rho^{AB} \rangle \sim (\rho^{AB})^{\otimes n}$. We will denote a *relative resource* as $\langle \mathcal{N} : \rho^A \rangle$ which is a channel guaranteed to behave as the channel \mathcal{N} provided the input state is exactly ρ^A .

As a first example of the protocol framework, consider the Schumacher compression result from the previous section. It can be represented by the following resource inequality

$$(H(B)_\sigma + \delta) [q \rightarrow q] \geq \langle \mathbf{1}^{A \rightarrow B} : \rho^A \rangle \quad (3.8)$$

for any $\delta \geq 0$ and where $\sigma^B := \mathbf{1}^{A \rightarrow B}(\rho^A)$. The above equation indicates that $(H(B) + \delta)$ qubits are sufficient to accurately convey the information contained in the state ρ^A to another party.

3.2.2 The family of quantum protocols

Many protocols of quantum information theory deal with the conversion of some noisy resource into the corresponding noiseless version possibly with the use of some auxiliary resources. It turns out that many of these protocols are related and it is sufficient to prove two protocols of this type and all other

protocols follow as simple consequences when we apply teleportation (TP) or superdense coding (SC) either before or after these protocols [3].

The two protocols which generate all the others of this “family tree” are called the mother and father protocols. The mother protocol takes the static resource $\langle \rho^{AB} \rangle$ and some quantum communication to distill maximally entangled bits. The resource inequality is

$$\langle \rho^{AB} \rangle + \frac{1}{2}I(A; R)_\psi[q \rightarrow q] \geq \frac{1}{2}I(A; B)_\psi[qq], \quad (\wp)$$

where the entropies are taken with respect to a purification $|\psi\rangle^{ABR}$ of ρ^{AB} . The mother protocol can be used to derive three “children”. The first of these is entanglement distillation, also known as the hashing inequality [44]. We start with equation (\wp) but implement the $[q \rightarrow q]$ term as teleportation

$$\frac{1}{2}I(A; R) \left([qq] + 2[c \rightarrow c] \geq [q \rightarrow q] \right). \quad (3.9)$$

After canceling some of the $[qq]$ terms on both sides we obtain

$$\langle \rho^{AB} \rangle + I(A; R)_\psi[c \rightarrow c] \geq I_c(A)B)_\psi[qq], \quad (3.10)$$

where $I_c(A)B) = \frac{1}{2}I(A; B) - \frac{1}{2}I(A; R) = H(B) - H(AB)$. The mother inequality can also be used to derive noisy versions of the teleportation [3] and superdense coding protocols [45].

The father protocol takes the dynamic resource of a noisy quantum channel $\langle \mathcal{N}^{A \rightarrow B} \rangle$ and some additional entanglement to simulate a noiseless quantum channel. Consider a setup where we send half of the state $|\phi\rangle^{AR}$ through the channel \mathcal{N} , which we model as an isometric extension $U_{\mathcal{N}}^{A \rightarrow BE}$ to an environment E . The resulting state is $|\psi\rangle^{BER} = U_{\mathcal{N}}^{A \rightarrow BE} \otimes \mathbf{1}^R |\phi\rangle^{AR}$ and the resource

inequality is

$$\langle \mathcal{N}^{A \rightarrow B} \rangle + \frac{1}{2} I(R; E)_\psi[qq] \geq \frac{1}{2} I(R; B)_\psi[q \rightarrow q]. \quad (\sigma)$$

Using the father inequality and the superdense coding result (SC), we can derive the formula for the entanglement-assisted classical capacity of a quantum channel [8]

$$\langle \mathcal{N}^{A \rightarrow B} : \rho^A \rangle + H(R)_\psi[qq] \geq I(R; B)_\psi[c \rightarrow c]. \quad (3.11)$$

More importantly, we can obtain the important quantum capacity result, the LSD Theorem [46, 47, 9], named after Lloyd, Shor and Devetak:

$$\langle \mathcal{N} \rangle \geq I_c(R; B)_\psi[q \rightarrow q]. \quad (3.12)$$

Furthermore, it turns out that equation (φ) and (σ) are related. We can obtain one from the other by replacing dynamic resources with static resources and adjusting for the definitions of A and R . This duality could be a mere coincidence or it could be indicative of some hidden structure. We will see in section 3.4 that in fact there exists an even bigger mother! The FQSW protocol, sometimes called “the mother of all protocols”, is a quantum protocol that generates both the mother and father protocols as well as many other protocols that were not part of the original family tree [5, 48].

3.3 State merging

Consider a setup where Alice and Bob share the state $\rho^{AB} = \text{Tr}_R |\psi\rangle\langle\psi|^{ABR}$. We would like to know how much quantum information Alice needs to send to Bob to *merge* her part of the state into Bob’s. The problem is illustrated graphically in Figure 3–1.

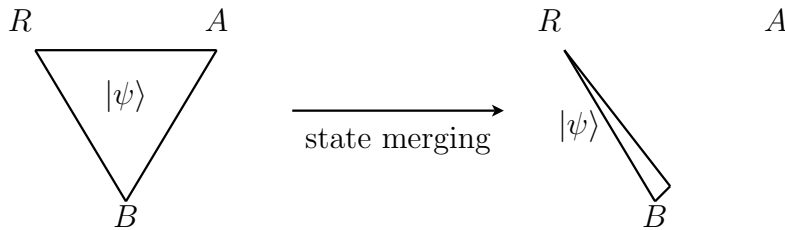


Figure 3–1: Pictorial representation of the state merging protocol. Alice’s part of $|\psi\rangle^{ABR}$ is merged with Bob’s part. In the end, the purification of R is held entirely in Bob’s system.

In the limit of many copies of the state, the rate at which Alice needs to send quantum information to Bob is given by the formula

$$R > H(A|B)_\rho, \quad (3.13)$$

provided classical communication is available for free. The primitive which optimally achieves this task is called the *state merging protocol* [15, 43]. We will discuss this protocol in some detail in section 3.3.2 but before that we dedicate some time to the quantum conditional entropy.

3.3.1 Quantum conditional entropy

The classical notion of conditional entropy $H(X|Y)$ is the amount of communication needed to convey the information content of the source X given knowledge of the variable Y at the decoder. As we saw in section 2.1.6, the conditional entropy is naturally suited to application in the Slepian-Wolf problem of distributed compression.

When we try to adapt the conditional entropy to the quantum world, we run into a number of conceptual difficulties. Indeed, in order to define $H(A|B)_\rho$ for a quantum state ρ^{AB} we need to replace the classical notion of a conditional distribution with some concept better suited to density matrices [49, 50]. A more pragmatic approach is to simply mimic the form of equation (2.16) from Section 2.1.5 and write the conditional entropy as a difference of two regular

entropies

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho. \quad (3.14)$$

In this way, we obtain a *formula* for the conditional entropy but still lack an *interpretation*. The situation is complicated further by the fact that the quantum conditional entropy can take on negative values seemingly indicating that it is possible to know more about the global state than about a part of it! Also to be explained is the relation between the negative values of the conditional entropy and the presence of quantum entanglement as indicated by the entropic Bell inequalities [51].

The interpretation issues around the quantum conditional entropy were finally settled in a satisfactory manner in two recent papers [15, 43], in which the quantum conditional entropy is given an operational interpretation in terms of the state merging protocol.

3.3.2 The state merging protocol

Consider a state ρ^{AB} shared between Alice and Bob and a purification of that state $|\psi\rangle^{ABR}$. We want to send Alice's part of the state to Bob by using an unlimited amount of classical communication and as little quantum communication as possible. Let $\Phi_K \in \mathcal{H}^{A_0 B_0}$, $\Phi_L \in \mathcal{H}^{A_1 B_1}$ be two maximally entangled states of rank K and L respectively. The state merging protocol takes as inputs the state $|\psi\rangle\langle\psi|^{ABR}$ and $\log K$ ebits in the form of Φ_K and applies the quantum operation $\mathcal{M}: AA_0 \otimes BB_0 \rightarrow A_1 \otimes B_1 \hat{B}B$ to produce a state

$$\sigma^{A_1 B_1 \hat{B} B R} = (\mathcal{M} \otimes \mathbf{1}^R)(|\psi\rangle\langle\psi|^{ABR} \otimes \Phi_K). \quad (3.15)$$

We want the state ρ^{AB} to be transferred entirely to Bob's lab: $\sigma^{\hat{B}B} \approx \rho^{AB}$. In addition, $\log L$ ebits are generated by the protocol if $\sigma^{A_1 B_1} \approx \Phi_L$. More

precisely, we measure the success of the protocol by the entanglement fidelity

$$F\left(\sigma^{A_1 B_1 \hat{B} B R}, \Phi_L^{A_1 B_1} \otimes |\psi\rangle\langle\psi|^{ABR}\right) \geq 1 - \epsilon. \quad (3.16)$$

In our original formulation of the state transfer task we asked how much quantum *communication* from Alice to Bob is necessary, yet in the above formulation we only speak of *entanglement* being consumed and generated. This is so because the two resources become equivalent when unlimited classical communication is allowed:

$$[qq] \equiv [q \rightarrow q] \quad (\text{free classical communication}). \quad (3.17)$$

Thus, we can say that Φ_K is the entanglement consumed by the protocol while Φ_L is the entanglement generated. In the i.i.d. regime, where $|\psi\rangle^{ABR} = \left(|\varphi\rangle^{ABR}\right)^{\otimes n}$, we define the *entanglement rate*

$$R = \frac{1}{n} (\log K - \log L), \quad (3.18)$$

which can take on both positive and negative values. When $R > 0$, the entanglement resource has been consumed by the protocol, but when $R < 0$ the protocol is actually generating entanglement as stated in the following theorem.

Theorem 3.3 (Quantum state merging [43]). *For a state ρ^{AB} shared by Alice and Bob, the entanglement cost of merging is equal to the quantum conditional entropy $H(A|B) = H(AB) - H(B)$. When $H(A|B)$ is positive, merging is possible only if $R > H(A|B)$ ebits per input copy are provided. When $H(A|B)$ is negative, the merging is possible by local operations and classical communication and moreover, $R < -H(A|B)$ maximally entangled states are obtained per input copy.*

We can express the state merging protocol as a resource inequality

$$\langle U^{S \rightarrow AB} : \rho^S \rangle + H(A|B)_\psi[q \rightarrow q] \geq \langle \mathbf{1}^{S \rightarrow B} : \rho^S \rangle \quad (\text{free } [c \leftrightarrow c]) \quad (3.19)$$

where $U^{S \rightarrow AB}$ is an isometry, $\rho^{AB} = U^{S \rightarrow AB}(\rho^S)$ that splits the state produced by the source between Alice & Bob while $\mathbf{1}^{S \rightarrow B}$ gives the state directly to Bob. The net effect of $\langle U^{S \rightarrow AB} : \rho^S \rangle$ on the left hand side and $\langle \mathbf{1}^{S \rightarrow B} : \rho^S \rangle$ on the right, is the state transfer resource informally defined

$$\langle \mathbf{1}^{A \rightarrow \hat{B}} : \rho^{AB} \rangle := \langle \mathbf{1}^{S \rightarrow B} : \rho^S \rangle - \langle U^{S \rightarrow AB} : \rho^S \rangle. \quad (3.20)$$

According to equation (3.19), this resource can be an asset or a liability depending on the sign of $H(A|B)$.

The state merging protocol has numerous applications. It can be used to study the quantum capacity of multiple access channels, entanglement distillation[11], entanglement of assistance[52] and distributed compression. The latter of these is of particular relevance to the subject of this thesis since it is a quantum generalization of the Slepian-Wolf problem discussed in section 2.1.6. Alice and Bob have to individually compress their shares of a state ρ^{AB} and transmit them to common receiver, Charlie. We allow unlimited classical communication and rates R_A, R_B of quantum communication to Charlie. The rate region for quantum distributed compression is given by the inequalities

$$\begin{aligned} R_A &> H(A|B)_\rho, \\ R_B &> H(B|A)_\rho, \\ R_A + R_B &> H(AB)_\rho. \end{aligned} \quad (3.21)$$

The rates for quantum distributed compression (3.21) should be compared with the classical distributed compression rates (2.20). This is an instance of

a general trend in quantum information theory: if classical communication is available for free, the solution to the quantum analogue of a given classical communication task is identical to the classical solution up to replacement of Shannon entropies by von Neumann entropies. Many times, however, this “ H goes to S rule” is only skin deep and sometimes it does not hold at all.

In the next section we will give the details of the fully quantum Slepian-Wolf (FQSW) protocol, which is a generalization of state merging where no classical communication is allowed. In the light of this, a detailed proof of the state merging protocol has been omitted for the sake of brevity and since it follows from the more powerful FQSW protocol.

3.4 The fully quantum Slepian-Wolf protocol

The fully quantum Slepian-Wolf protocol [5] is a procedure for simultaneous quantum state transfer and entanglement distillation. It can be thought of as the quantum version of the classical Slepian-Wolf protocol but, unlike the state merging protocol considered above, no classical communication is allowed. This FQSW protocol generates nearly all the other protocols of quantum information theory as special cases, yet despite its powerful applications it is fairly simple to implement.

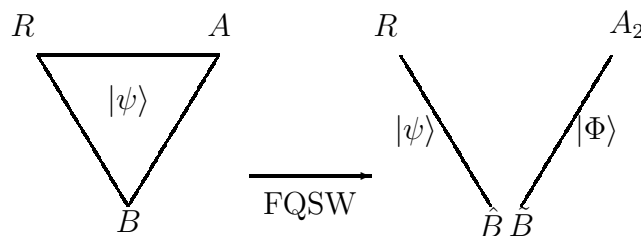


Figure 3–2: Diagram representing the ABR correlations before and after the FQSW protocol. Alice manages to decouple completely from the reference R . The \hat{B} system is isomorphic to the original AB : it is the purification of R .

The state $|\psi\rangle^{ABR} = \left(|\varphi\rangle^{ABR}\right)^{\otimes n}$ is shared between Alice, Bob and a reference system R . The FQSW protocol describes a procedure for Alice to

transfer her R -entanglement to Bob while at the same time generating ebits with him. Alice can accomplish this by encoding and sending part of her system, denoted A_1 , to Bob. The state after the protocol can approximately be written as $|\Phi\rangle^{A_2\tilde{B}}(|\varphi\rangle^{R\hat{B}})^{\otimes n}$, where the systems \tilde{B} and \hat{B} are held in Bob's lab while A_2 remains with Alice. The additional product, $|\Phi\rangle^{A_2\tilde{B}}$, is a maximally entangled state shared between Alice and Bob. Figure 3–2 illustrates the entanglement structure before and after the protocol.

3.4.1 The protocol

The protocol relies on an initial compression step and the mixing effect of random unitary operations for the encoding. We assume that, before the start of the protocol, Alice and Bob have pre-chosen a random unitary operation U_A . Equivalently, they could have shared random bits which they use to locally generate the same unitary operation.

The protocol, represented graphically in Figure 3–3, consists of the following steps:

1. Alice performs Schumacher compression on her system A to obtain the output system A^S .
2. Alice then applies a random unitary U_A to A^S .
3. Next, she splits her system into two parts: $A_1A_2 = A^S$ with $d_{A_1} = 2^{nQ_A}$ and

$$Q_A > \frac{1}{2}I(A; R)_\varphi. \quad (3.22)$$

She sends the system A_1 to Bob.

4. Bob, in turn, performs a decoding operation $V_B^{A_1B \rightarrow \hat{B}\tilde{B}}$ which splits his system into a \hat{B} part purifying R and a \tilde{B} part which is fully entangled with Alice.

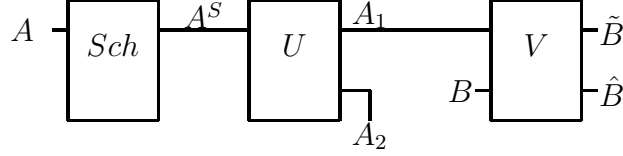


Figure 3–3: Circuit diagram for the FQSW protocol. First we Schumacher compress the A system, then we apply the random unitary encoding U_A . At the receiving end Bob applies a decoding operation V .

The best way to understand the mechanism behind this protocol is by thinking about destroying correlations. If, at the end of the protocol, Alice's system A_2 is nearly decoupled from the reference in the sense that $\sigma^{A_2 R} \approx \sigma^{A_2} \otimes \sigma^R$, then Alice must have succeeded in sending her R entanglement to Bob because it is Bob alone who then holds the R purification. We can therefore guess the lower bound on how many qubits Alice will have to send before she can decouple from the reference. Originally, Alice and R share $I(A; R)_\varphi$ bits of information per copy of $|\varphi\rangle^{ABR}$. Since one qubit can carry away at most two bits of quantum mutual information, this means that the minimum rate at which Alice must send qubits to Bob is

$$Q_A > \frac{1}{2} I(A; R)_\varphi. \quad (3.23)$$

It is shown in [5] that this rate is achievable in the limit of many copies of the state. Therefore the FQSW protocol is optimal for the state transfer task. More formally the decoupling process is described by the following theorem:

Theorem 3.4 (One-shot decoupling theorem from [5]).

Let $\sigma^{A_2 R}(U) = \text{Tr}_{A_1}[(U \otimes \mathbf{1}^R) \psi^{A^S R} (U^\dagger \otimes \mathbf{1}^R)]$ be the state remaining on $A_2 R$ after the unitary transformation U has been applied to $A^S = A_1 A_2$. Then

$$\int_{\mathbb{U}(A)} \left\| \sigma^{A_2 R}(U) - \frac{\mathbf{1}^{A_2}}{d_{A_2}} \otimes \sigma^R \right\|_1^2 dU \leq \frac{d_{A^S} d_R}{d_{A_1}^2} \text{Tr}[(\psi^{A^S R})^2]. \quad (3.24)$$

This theorem quantifies how close to decoupled the A_2 and R systems are if a random unitary operation is applied to the $A^S = A_1 A_2$ system. There are several important observations to make in relation to the above inequality. First, we note that for a given state $|\psi\rangle^{ABR}$, the dimensions of the systems A^S and R as well as the purity $\text{Tr}[(\psi^{A^S R})^2]$ are fixed numbers over which Alice has no control. Alice can, however, choose the dimension of the subsystem she sends to Bob, d_{A_1} , and influence how decoupled she is from the reference. By making making d_{A_1} sufficiently large, Alice can thus make the right hand side of (3.24) tend to zero.

Second, the fact that Alice holds something very close to a maximally mixed state $\mathbf{1}/d_{A_2}$ indicates that Bob can, by an appropriate choice of decoding operation V_B , establish a maximally entangled state $|\Phi\rangle^{A_2 \tilde{B}}$ with Alice. These ebits generated between Alice and Bob are a useful side-effect of the protocol that is similar to the entanglement generated by the state merging protocol.

All that remains now is to specify d_{A_1} , the dimension of the system sent to Bob, in terms of entropic quantities of the input state. This can be done in the the limit where n , the number of copies of the state goes to infinity. Using the properties of typical subspaces, we can we can make the right hand side of equation (3.24) tend to zero provided the rate $Q_A \equiv \frac{1}{n} \log d_{A_1}$ satisfies [5]:

$$Q_A \geq \frac{1}{2} I(A; R)_\varphi + \delta \quad (3.25)$$

for any $\delta > 0$.

3.4.2 The FQSW resource inequality

In the spirit of section 3.2 above, we can succinctly express the effects of the fully-quantum Slepian-Wolf protocol as a resource inequality

$$\langle U^{S \rightarrow AB} : \varphi^S \rangle + \frac{1}{2} I(A; R)_\varphi [q \rightarrow q] \geq \frac{1}{2} I(A; B)_\varphi [qq] + \langle \mathbf{1}^{S \rightarrow \hat{B}} : \varphi^S \rangle \quad (3.26)$$

which is read: *Given the state $|\varphi\rangle^{ABR}$ and $\frac{1}{2}I(A;R)$ qubits of communication from Alice to Bob we can obtain the state $|\varphi\rangle^{R\hat{B}}$ while also purifying $\frac{1}{2}I(A;B)$ ebits.*

As previously announced, the FQSW protocol is more powerful than the state merging protocol of section 3.3 since it generates it as a special case. Indeed, when we implement the quantum communication $[q \rightarrow q]$ of equation (3.26) as teleportation according to equation (TP)

$$\frac{1}{2}I(A;R)[qq] + I(A;R)[c \rightarrow c] \geq \frac{1}{2}I(A;R)[q \rightarrow q]. \quad (3.27)$$

We now “recycle” the entanglement produced by the protocol. The factor in front of $[qq]$ is going to be $\frac{1}{2}I(A;R) - \frac{1}{2}I(A;B) = H(A|B)$ and the overall resource inequality becomes

$$\langle U^{S \rightarrow AB} : \varphi^S \rangle + H(A|B)_\varphi[qq] + I(A;R)_\varphi[c \rightarrow c] \geq \langle \mathbf{1}^{S \rightarrow \hat{B}} : \varphi^S \rangle, \quad (3.28)$$

which is exactly the state merging resource inequality (3.19), when we also account for the classical communication cost.

The FQSW inequality generates the mother inequality (\wp) by discarding the additional resource $\langle \mathbf{1}^{S \rightarrow \hat{B}} : \varphi^S \rangle$ on the right hand side. Moreover it was recently shown that by the *source-channel duality*, the FQSW protocol can be used to generate the father protocol (σ) and by *time reversal duality* the FQSW protocol leads to the fully quantum reverse Shannon (FQRS) protocol [48]. Other notable results related to the FQSW protocol are the recent results for broadcast channels [53], and the generalization of the FQSW task called *quantum state redistribution*, which uses side information both at the encoder and the decoder [54, 55].

In addition to the its powerful protocol generating faculties, the FQSW protocol has applications to the distributed compression problem for quantum systems. Indeed, the original FQSW paper [5] partially solves the distributed compression problem in the two-party case by providing upper and lower bounds on the set of achievable rates. In Chapter 5 we will present our results on the multiparty version of the same problem. For the sake of continuity, the reader may wish to skip Chapter 4 on a first reading of the thesis since it is a self-contained exposition on the multiparty squashed entanglement, which only comes into play relatively late in the distributed compression chapter.

CHAPTER 4

Multiparty quantum information

Many of the protocols of information theory deal with multiple senders and multiple receivers. As a whole, however, *network information theory*, the field which studies general multiparty communication scenarios is not yet fully developed even for classical systems [26]. Quantum network information theory, which deals with quantum multipartite communication, is also under active development [56, 57, 25] and, thanks to the no-cloning properties of quantum information, sometimes admits simple solutions [56]. On the other hand, a full understanding of quantum network theory will require a precise characterization of multiparty entanglement, a task which is far from completed [21, 22, 23, 24]. Nevertheless, we can hope that years from now we will have a rigorous and complete theory of multiparty information theory in the spirit of the two-party protocols framework [4].

One step toward the development of a multiparty information theory would be to generalize the concept of mutual information $I(A; B)$ to more than two parties. The mutual information, the information that two systems A and B have in common, can be written as

$$I(A; B) = H(A) - H(A|B). \tag{4.1}$$

The above formula is interpreted as a reduction of the total uncertainty of A by the amount that is not common to B . What is left is the uncertainty that is shared.

Another way to write the mutual information is

$$I(A; B) = H(A) + H(B) - H(AB), \quad (4.2)$$

which adds both entropies (double-counting the entropy that is common) and then subtracts the total entropy. Equation (4.2) is a measure of how different from independent the variables A and B are. Both of these interpretations of the mutual information can be generalized to the multiparty case.

One way to define the multiparty mutual information for three variables A, B and C is by mimicking the form of equation (4.1) above and define

$$I_{\cap}(A; B; C) := I(A; B) - I(A; B|C). \quad (4.3)$$

The motivation behind this formula is to subtract from the mutual information $I(A; B)$ any terms that are due to AB -only correlations and not true tripartite correlations. The expanded form of the restrictive mutual information is

$$I_{\cap}(A; B; C) = H(A) + H(B) + H(C) - H(AB) - H(AC) - H(BC) + H(ABC).$$

This is the information shared by *all* three parties and corresponds to the region labeled “g” in Figure 4–1. This form of multiparty mutual information was defined in [50] but has not yet proved useful in applications. Also, $I_{\cap}(A; B; C)$ can take on negative values [16], which are difficult to interpret.

Another approach is to define the multiparty mutual information in the spirit of (4.2), as the measure of how different from independent the three variables are

$$I_{\cup}(A; B; C) := H(A) + H(B) + H(C) - H(ABC). \quad (4.4)$$

In terms of the regions in Figure 4–1, we have $I_{\cup}(A; B; C) = d + e + f + 2g$. This form of the mutual information is naturally connected to the relative

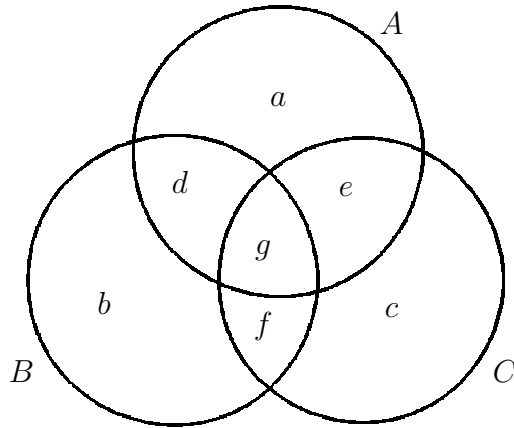


Figure 4–1: Entropy diagram for three parties A , B and C .

entropy[27] and also satisfies the chain-like property

$$I_{\cup}(A; B; C) = I(A; B) + I(AB; C), \quad (4.5)$$

which indicates how the multiparty information is affected when we introduce a new system.

In this chapter we will investigate some of the properties of the inclusive multiparty information $I_{\cup}(A; B; C)$, henceforth referred to simply as mutual information $I(A; B; C)$. Our work on the multiparty information will also allow us to naturally extend the notion of squashed entanglement [17] to the multiparty scenario. The multiparty squashed entanglement, discussed in Section 4.2, turns out to be a measure of multipartite entanglement with excellent properties and clear and intuitive interpretation. It finds application in the proof of Theorem 5.3, the outer bound on the rate region for distributed compression.

4.1 Multiparty information

We begin with the definition of the multiparty quantum information for m parties.

Definition 4.1 (Multiparty information). *Given the state $\rho^{X_1 X_2 \dots X_m}$ shared between m systems, we define the multiparty information as:*

$$\begin{aligned} I(X_1; X_2; \dots; X_m)_\rho &:= H(X_1) + H(X_2) + \dots + H(X_m) - H(X_1 X_2 \dots X_m) \\ &= \sum_{i=1}^m H(X_i)_\rho - H(X_1 X_2 \dots X_m)_\rho \end{aligned} \quad (4.6)$$

The subadditivity inequality for quantum entropy ensures that the multiparty information is zero if and only if ρ has the tensor product form $\rho^{X_1} \otimes \rho^{X_2} \otimes \dots \otimes \rho^{X_m}$.

The conditional version of the multiparty mutual information is obtained by replacing all the entropies by conditional entropies

$$\begin{aligned} I(X_1; X_2; \dots; X_m|E)_\rho &:= \sum_{i=1}^m H(X_i|E) - H(X_1 X_2 \dots X_m|E) \\ &= \sum_{i=1}^m H(X_i E) - H(X_1 X_2 \dots X_m E) - (m-1)H(E) \\ &= I(X_1; X_2; \dots; X_m; E) - \sum_{i=1}^m I(X_i; E). \end{aligned} \quad (4.7)$$

This definition of multiparty information has appeared previously in [58, 59, 60] and more recently in [16], where many of its properties were investigated.

Next we investigate some formal properties of the multiparty information which will be useful in our later analysis.

Lemma 4.2 (Merging of multiparty information terms). *Arguments of the multiparty information can be combined by subtracting their mutual information:*

$$I(A; B; X_1; X_2; \dots; X_m) - I(A; B) = I(AB; X_1; X_2; \dots; X_m). \quad (4.8)$$

Proof. This identity is a simple calculation. It is sufficient to expand the definitions and cancel terms.

$$\begin{aligned}
I(A; B; X_1; X_2; \dots; X_m) - I(A; B) &= \\
&= H(A) + H(B) + \sum H(X_i) - H(ABX_1X_2 \dots X_m) - H(A) - H(B) + H(AB) \\
&= H(AB) + \sum H(X_i) - H(A, BX_1X_2 \dots X_m) \\
&= I(AB; X_1; X_2; \dots; X_m).
\end{aligned}$$

□

Discarding a subsystem inside the conditional multiparty information cannot lead it to increase. This property, more than any other, justifies its use as a measure of correlation.

Lemma 4.3 (Monotonicity of conditional multiparty information).

$$I(AB; X_1; \dots X_m | E) \geq I(A; X_1; \dots X_m | E) \quad (4.9)$$

Proof. This follows easily from strong subadditivity of quantum entropy (SSA).

$$\begin{aligned}
I(AB; X_1; X_2; \dots; X_m | E) &= \\
&= H(ABE) + \sum_i H(X_i E) - H(ABX_1X_2 \dots X_mE) - mH(E) \\
&= H(ABE) + \sum_i H(X_i E) - H(ABX_1X_2 \dots X_mE) - mH(E) + \\
&\quad \underbrace{H(AE) - H(AE)}_{=0} + \underbrace{H(AX_1X_2 \dots X_mE) - H(AX_1X_2 \dots X_mE)}_{=0} \\
&= H(AE) + \sum_i H(X_i E) - H(AX_1X_2 \dots X_mE) - mH(E) + \\
&\quad \underbrace{[H(ABE) + H(AX_1X_2 \dots X_mE) - H(AE) - H(ABX_1X_2 \dots X_mE)]}_{\geq 0 \text{ by SSA}}
\end{aligned}$$

$$\begin{aligned}
&\geq H(AE) + \sum_i H(X_i E) - H(AX_1 X_2 \dots X_m E) - mH(E) \\
&= I(A; X_1; X_2 \dots X_m | E)
\end{aligned}$$

□

We will now prove a multiparty information property that follows from a more general chain rule, but is all that we will need for applications.

Lemma 4.4 (Chain-type Rule).

$$I(AA'; X_1; \dots; X_m | E) \geq I(A; X_1; \dots; X_m | A'E) \quad (4.10)$$

Proof.

$$\begin{aligned}
I(AA'; X_1; \dots; X_m | E) &= \\
&= H(AA'E) + \sum_{i=1}^m H(X_i E) - H(AA'X_1, \dots, X_m) - mH(E) \\
&= I(A; X_1; \dots; X_m | A'E) + \sum_{i=1}^m [H(A'E) + H(X_i E) - H(E) - H(A'X_i E)] \\
&\geq I(A; X_1; \dots; X_m | A'E).
\end{aligned}$$

The inequality is true by strong subadditivity. □

Remark It is interesting to note that we have two very similar reduction-of-systems formulas derived from different perspectives. From Lemma 4.3 (monotonicity of the multiparty information) we have that

$$I(AB; X_1; \dots; X_m | E) \geq I(A; X_1; \dots; X_m | E), \quad (4.11)$$

but we also know from Lemma 4.4 (chain-type rule) that

$$I(AB; X_1; \dots; X_m | E) \geq I(A; X_1; \dots; X_m | BE). \quad (4.12)$$

The two expressions are inequivalent; one is not strictly stronger than the other. We use both of them depending on whether we want to keep the deleted system around for conditioning.

4.2 Multiparty squashed entanglement

Using the definition of the conditional multiparty information from the previous section, we can define a multiparty squashed entanglement analogous to the bipartite version [61, 62, 17]. The multiparty squashed entanglement has recently been investigated independently by Yang et al. [16].

Definition 4.5 (Multiparty squashed entanglement). *Consider the density matrix $\rho^{X_1 X_2 \dots X_m}$ shared between m parties. We define the multiparty squashed entanglement in the following manner*

$$\begin{aligned} E_{sq}(X_1; X_2; \dots; X_m)_\rho &:= \frac{1}{2} \inf_E \left[\sum_{i=1}^m H(X_i|E)_{\tilde{\rho}} - H(X_1 X_2 \dots X_m|E)_{\tilde{\rho}} \right] \\ &= \frac{1}{2} \inf_E I(X_1; X_2; \dots; X_m|E)_{\tilde{\rho}} \end{aligned} \quad (4.13)$$

where the minimization happens over all states of the form $\tilde{\rho}^{X_1 X_2 \dots X_m E}$ such that $\text{Tr}_E(\tilde{\rho}^{X_1 X_2 \dots X_m E}) = \rho^{X_1 X_2 \dots X_m}$. (We say $\tilde{\rho}$ is an extension of ρ .)

The dimension of the extension system E can be arbitrarily large, which is in part what makes calculations of the squashed entanglement very difficult except for simple systems. The motivation behind this definition is that we can include a copy of all classical correlations inside the extension E and thereby eliminate them from the multiparty information by conditioning. Since it is impossible to copy quantum information, we know that taking the infimum over all possible extensions E we will be left with a measure of the purely quantum correlations. The definition of E_{sq} as a minimization over a conditional mutual information is motivated by the classical cryptography notion of *intrinsic information* which provides a bound on the secret-key rate [63, 64, 17].

Example: It is illustrative to calculate the squashed entanglement for separable states, which are probabilistic mixtures of tensor products of local pure states. Consider the state

$$\rho^{X_1 X_2 \dots X_m} = \sum_j p_j |\alpha_j\rangle\langle\alpha_j|^{X_1} \otimes |\beta_j\rangle\langle\beta_j|^{X_2} \otimes \dots \otimes |\zeta_j\rangle\langle\zeta_j|^{X_m},$$

which we choose to extend by adding a system E containing a record of the index j as follows

$$\tilde{\rho}^{X_1 X_2 \dots X_m E} = \sum_j p_j |\alpha_j\rangle\langle\alpha_j|^{X_1} \otimes |\beta_j\rangle\langle\beta_j|^{X_2} \otimes \dots \otimes |\zeta_j\rangle\langle\zeta_j|^{X_m} \otimes |j\rangle\langle j|^E.$$

When we calculate conditional entropies we notice that for any subset $\mathcal{K} \subseteq \{1, 2, \dots, m\}$,

$$H(X_{\mathcal{K}}|E)_{\tilde{\rho}} = 0. \quad (4.14)$$

Knowledge of the classical index leaves us with a pure product state for which all the relevant entropies are zero. Therefore, separable states have zero squashed entanglement:

$$E_{\text{sq}}(X_1; X_2; \dots; X_m)_{\rho} = \frac{1}{2} \left[\sum_i^m H(X_i|E)_{\tilde{\rho}} - H(X_1 X_2 \dots X_m | E)_{\tilde{\rho}} \right] = 0.$$

We now turn our attention to the properties of E_{sq} . Earlier we argued that the squashed entanglement measures purely quantum contributions to the mutual information between systems, in the sense that it is zero for all separable states. In this section we will show that the multiparty squashed entanglement cannot increase under the action of local operations and classical communication, that is, that E_{sq} is an LOCC-monotone. We will also show that E_{sq} has other desirable properties; it is convex, subadditive and continuous.

Proposition 4.6. *The quantity E_{sq} is an entanglement monotone, i.e. it does not increase on average under local quantum operations and classical communication (LOCC).*

Proof. In order to show this we will follow the argument of [17], which in turn follows the approach described in [65]. We will show that E_{sq} has the following two properties:

1. Given any unilocal quantum instrument \mathcal{E}_k (a collection of completely positive maps such that $\sum_k \mathcal{E}_k$ is trace preserving [66]) and any quantum state $\rho^{X_1 \dots X_m}$, then

$$E_{sq}(X_1; X_2; \dots X_m)_\rho \geq \sum_k p_k E_{sq}(X_1; X_2; \dots X_m)_{\tilde{\rho}_k} \quad (4.15)$$

where

$$p_k = \text{Tr } \mathcal{E}_k(\rho^{X_1 \dots X_m}) \quad \text{and} \quad \tilde{\rho}_k^{X_1 \dots X_m} = \frac{1}{p_k} \mathcal{E}_k(\rho^{X_1 \dots X_m}). \quad (4.16)$$

2. E_{sq} is convex.

Without loss of generality, we assume that \mathcal{E}_k acts on the first system. We will implement the quantum instrument by appending to X_1 environment systems X'_1 and X''_1 prepared in standard pure states, applying a unitary U on $X_1 X'_1 X''_1$, and then tracing out over X''_1 . We store k , the classical record of which \mathcal{E}_k occurred, in the X'_1 system. More precisely, for any extension of $\rho^{X_1 X_2 \dots X_m}$ to $X_1 X_2 \dots X_m E$,

$$\rho^{X_1 X_2 \dots X_m E} \mapsto \tilde{\rho}^{X_1 X'_1 X_2 \dots X_m E} := \sum_k \mathcal{E}_k \otimes I_E(\rho^{X_1 X_2 \dots X_m E}) \otimes |k\rangle \langle k|^{X'_1}. \quad (4.17)$$

The argument is then as follows:

$$\frac{1}{2}I(X_1; X_2; \dots X_m|E)_\rho = \frac{1}{2}I(X_1X'_1X''_1; X_2; \dots; X_m|E)_\rho \quad (4.18)$$

$$= \frac{1}{2}I(X_1X'_1X''_1; X_2; \dots; X_m|E)_{\tilde{\rho}} \quad (4.19)$$

$$\geq \frac{1}{2}I(X_1X'_1; X_2; \dots; X_m|E)_{\tilde{\rho}} \quad (4.20)$$

$$\geq \frac{1}{2}I(X_1; X_2; \dots; X_m|EX'_1)_{\tilde{\rho}} \quad (4.21)$$

$$= \frac{1}{2} \sum_k p_k I(X_1; X_2; \dots; X_m|E)_{\tilde{\rho}_k} \quad (4.22)$$

$$\geq \sum_k p_k E_{\text{sq}}(X_1; X_2; \dots; X_m)_{\tilde{\rho}_k} \quad (4.23)$$

The equality (4.18) is true because adding an uncorrelated ancilla does not change the entropy of the system. The transition $\rho \rightarrow \tilde{\rho}$ is unitary and doesn't change entropic quantities so (4.19) is true. For (4.20) we use the monotonicity of conditional multiparty information, Lemma 4.3. In (4.21) we use the chain-type rule from Lemma 4.4. In (4.22) we use the index information k contained in X'_1 . Finally, since E_{sq} is the infimum over all extensions, it must be no more than the particular extension E , so (4.23) must be true. Now since the extension E in (4.18) was arbitrary, it follows that $E_{\text{sq}}(X_1; X_2; \dots; X_m)_\rho \geq \sum_k p_k E_{\text{sq}}(X_1; X_2; \dots; X_m)_{\tilde{\rho}_k}$ which completes the proof of Property 1.

To show the convexity of E_{sq} , we again follow the same route as in [17]. Consider the states $\rho^{X_1X_2\dots X_m}$ and $\sigma^{X_1X_2\dots X_m}$ and their extensions $\tilde{\rho}^{X_1X_2\dots X_mE}$ and $\tilde{\sigma}^{X_1X_2\dots X_mE}$ defined over the same system E . We can also define the weighted sum of the two states $\tau^{X_1X_2\dots X_m} = \lambda\rho^{X_1X_2\dots X_m} + (1-\lambda)\sigma^{X_1X_2\dots X_m}$ and the following valid extension:

$$\tilde{\tau}^{X_1X_2\dots X_mE E'} = \lambda\rho^{X_1X_2\dots X_mE} \otimes |0\rangle\langle 0|^{E'} + (1-\lambda)\sigma^{X_1X_2\dots X_mE} \otimes |1\rangle\langle 1|^{E'}. \quad (4.24)$$

Using the definition of squashed entanglement we know that

$$\begin{aligned}
E_{\text{sq}}(X_1; X_2; \dots; X_m)_\tau & \\
&\leq \frac{1}{2} I(X_1; X_2; \dots; X_m | EE')_{\tilde{\tau}} \\
&= \frac{1}{2} [\lambda I(X_1; X_2; \dots; X_m | E)_{\tilde{\rho}} + (1 - \lambda) I(X_1; X_2; \dots; X_m | E)_{\tilde{\sigma}}].
\end{aligned}$$

Since the extension system E is completely arbitrary we have

$$E_{\text{sq}}(X_1; \dots; X_m)_\tau \leq \lambda E_{\text{sq}}(X_1; \dots; X_m)_\rho + (1 - \lambda) E_{\text{sq}}(X_1; \dots; X_m)_\sigma,$$

so E_{sq} is convex.

We have shown that E_{sq} satisfies both Properties 1 and 2 from page 51.

Therefore, it must be an entanglement monotone. \square

Subadditivity on Product States Another desirable property for measures of entanglement is that they should be additive or at least subadditive on tensor products of the same state. Subadditivity of E_{sq} is easily shown from the properties of multiparty information.

Proposition 4.7. *E_{sq} is subadditive on tensor product states, i.e.*

$$E_{\text{sq}}(X_1 Y_1; \dots; X_m Y_m)_\rho \leq E_{\text{sq}}(X_1; \dots; X_m)_\rho + E_{\text{sq}}(Y_1; \dots; Y_m)_\rho \quad (4.25)$$

where $\rho^{X_1 Y_1 X_2 Y_2 \dots X_m Y_m} = \rho^{X_1 X_2 \dots X_m} \otimes \rho^{Y_1 Y_2 \dots Y_m}$.

Proof. Assume that $\rho^{X_1 X_2 \dots X_m E}$ and $\rho^{Y_1 Y_2 \dots Y_m E'}$ are extensions. Together they form an extension $\rho^{X_1 Y_1 X_2 Y_2 \dots X_m Y_m EE'}$ for the product state.

$$\begin{aligned}
2E_{\text{sq}}(X_1 Y_1; X_2 Y_2; \dots; X_m Y_m)_\rho & \\
&\leq I(X_1 Y_1; X_2 Y_2; \dots; X_m Y_m | EE') \\
&= \sum_i H(X_i Y_i | EE') - H(X_1 Y_1 X_2 Y_2 \dots X_m Y_m | EE') - (m - 1)H(EE')
\end{aligned}$$

$$= I(X_1; X_2; \dots; X_m|E) + I(Y_1; Y_2; \dots; Y_m|E'). \quad (4.26)$$

The first line holds because the extension for the XY system that can be built by combining the X and Y extensions is not the most general extension. The proposition then follows because the inequality holds for all extensions of ρ and σ . \square

The question of whether E_{sq} is additive, meaning superadditive in addition to subadditive, remains an open problem. Indeed, if it were possible to show that correlation between the X and Y extensions is unnecessary in the evaluation of the squashed entanglement of $\rho \otimes \sigma$, then E_{sq} would be additive. This is provably true in the bipartite case [17] but the same method does not seem to work with three or more parties.

Continuity The continuity of bipartite E_{sq} was conjectured in [17] and proved by Alicki and Fannes in [67]. We will follow the same argument here to prove the continuity of the multiparty squashed entanglement. The key to the continuity proof is the following lemma which makes use of an ingenious geometric construction.

Lemma 4.8 (Continuity of conditional entropy [67]). *Given density matrices ρ^{AB} and σ^{AB} on the space $\mathcal{H}^A \otimes \mathcal{H}^B$ such that*

$$\|\rho - \sigma\|_1 = \frac{1}{2} \text{Tr}|\rho - \sigma| \leq \epsilon, \quad (4.27)$$

it is true that

$$|H(A|B)_\rho - H(A|B)_\sigma| \leq 4\epsilon \log d_A + 2h(\epsilon) \quad (4.28)$$

where $d_A = \dim \mathcal{H}^A$ and $h(\epsilon) = -\epsilon \log \epsilon - (1 - \epsilon) \log(1 - \epsilon)$ is the binary entropy.

This seemingly innocuous technical lemma makes it possible to prove the continuity of E_{sq} in spite of the unbounded dimension of the extension system.

Proposition 4.9 (E_{sq} is continuous). *For all states $\rho^{X_1 X_2 \dots X_m}$, $\sigma^{X_1 X_2 \dots X_m}$ with $\|\rho - \sigma\|_1 \leq \epsilon$, $\|E_{\text{sq}}(\rho) - E_{\text{sq}}(\sigma)\| \leq \epsilon'$ where ϵ' depends on ϵ and vanishes as $\epsilon \rightarrow 0$.*

The precise form of ϵ' can be found in equation (4.35).

Proof. Proximity in trace distance implies proximity in fidelity distance [40], in the sense that

$$F(\rho^{X_1 X_2 \dots X_m}, \sigma^{X_1 X_2 \dots X_m}) \geq 1 - \epsilon, \quad (4.29)$$

but by Uhlmann's theorem [68] this means that we can find purifications $|\rho\rangle^{X_1 X_2 \dots X_m R}$ and $|\sigma\rangle^{X_1 X_2 \dots X_m R}$ such that

$$F(|\rho\rangle^{X_1 X_2 \dots X_m R}, |\sigma\rangle^{X_1 X_2 \dots X_m R}) \geq 1 - \epsilon. \quad (4.30)$$

Now if we imagine some general operation Λ that acts only on the purifying system R

$$\rho^{X_1 X_2 \dots X_m E} = (I^{X_1 X_2 \dots X_m} \otimes \Lambda^{R \rightarrow E}) |\rho\rangle\langle\rho|^{X_1 X_2 \dots X_m R} \quad (4.31)$$

$$\sigma^{X_1 X_2 \dots X_m E} = (I^{X_1 X_2 \dots X_m} \otimes \Lambda^{R \rightarrow E}) |\sigma\rangle\langle\sigma|^{X_1 X_2 \dots X_m R} \quad (4.32)$$

we have from the monotonicity of fidelity for quantum channels that

$$F(\rho^{X_1 X_2 \dots X_m E}, \sigma^{X_1 X_2 \dots X_m E}) \geq F(|\rho\rangle^{X_1 X_2 \dots X_m R}, |\sigma\rangle^{X_1 X_2 \dots X_m R}) \geq 1 - \epsilon, \quad (4.33)$$

which in turn implies [40] that

$$\|\rho^{X_1 X_2 \dots X_m E} - \sigma^{X_1 X_2 \dots X_m E}\|_1 \leq 2\sqrt{\epsilon}. \quad (4.34)$$

Now we can apply Lemma 4.8 to each term in the multiparty information to obtain

$$\begin{aligned}
& \left| I(X_1; X_2; \dots X_m | E)_\rho - I(X_1; X_2; \dots X_m | E)_\sigma \right| \\
& \leq \sum_{i=1}^m \left| H(X_i | E)_\rho - H(X_i | E)_\sigma \right| \\
& \quad + \left| H(X_1 X_2 \dots X_m | E)_\rho - H(X_1 X_2 \dots X_m | E)_\sigma \right| \\
& \leq \sum_{i=1}^m [8\sqrt{\epsilon} \log d_i + 2h(2\sqrt{\epsilon})] + 8\sqrt{\epsilon} \log \left(\prod_{i=1}^m d_i \right) + 2h(2\sqrt{\epsilon}) \\
& = 16\sqrt{\epsilon} \log \left(\prod_{i=1}^m d_i \right) + (m+1)2h(2\sqrt{\epsilon}) =: \epsilon' \tag{4.35}
\end{aligned}$$

where $d_i = \dim \mathcal{H}^{X_i}$ and $h(\cdot)$ is as defined in Lemma 4.8. Since we have shown the above inequalities for *any* extension E and the quantity ϵ' vanishes as $\epsilon \rightarrow 0$, we have proved that E_{sq} is continuous. \square

4.3 Example calculations of E_{sq}

Below we give several examples of simple systems where E_{sq} is calculated to gain intuition about how it behaves. As a first step we verify that E_{sq} is zero for states that are manifestly not entangled.

Example 1: Fully decoupled state Given the state $\rho_1^{X_1 X_2 \dots X_m} = \rho^{X_1} \otimes \rho^{X_2} \otimes \dots \otimes \rho^{X_m} = \bigotimes_1^m \rho^{X_i}$ the mutual information for this state is:

$$I(X_1; X_2; \dots; X_m) = \sum_i^m H(X_i) - \underbrace{H(X_1, X_2, \dots, X_m)}_{=\sum_i^m H(X_i)} = 0 \tag{4.36}$$

which is to be expected since the state is a tensor product and cannot contain entanglement.

In the next example we look at more complicated states where E_{sq} is non-zero but simple to calculate.

Example 2: Partially separable state Now consider a state which is separable on all systems except for two. We write

$$\rho_2^{ABX_1X_2\cdots X_m} = \sum_j p_j |\alpha_j\rangle\langle\alpha_j|^{AB} \otimes |\beta_j\rangle\langle\beta_j|^{X_1} \otimes \cdots \otimes |\zeta_j\rangle\langle\zeta_j|^{X_m}$$

and an extension E that records the index j . For this extension we will have:

$$\begin{aligned} I(A; B; X_1; \dots; X_m | E) &= \\ &= H(A|E) + H(B|E) + \sum_i^m H(X_i|E) - H(ABX_1 \cdots X_m | E) \\ &= H(AE) + H(BE) + \sum_i^m H(X_i E) - H(ABX_1 \cdots X_m E) - (m+1)H(E) \\ &\geq^{(1)} H(AE) + H(BE) - H(ABE) - H(E) \\ &= I(A; B | E) \\ &\geq 2E_{\text{sq}}(A; B)_\rho \end{aligned}$$

To show (1) we repeatedly used the strong subadditivity property of von Neumann entropy

$$-H(E) - H(EY_1 \cdots Y_m) \geq -H(EY_m) - H(EY_1 \cdots Y_{m-1}). \quad (4.37)$$

Thus we have shown that for partially separable states, E_{sq} of the whole is at least as much as its non-separable part.

Example 3: E_{sq} for the GHZ and W states Consider the m -party GHZ state $|GHZ\rangle^{X_1 X_2 \cdots X_m} = \frac{|0\rangle^{\otimes m} + |1\rangle^{\otimes m}}{\sqrt{2}}$ and the m -party W state $|W\rangle^{X_1 X_2 \cdots X_m} = \frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} |\hat{i}\rangle$, where $|\hat{i}\rangle = |0 \cdots 0 1 \underbrace{0 \cdots 0}_i\rangle$. In particular, the three-party GHZ and W states correspond to

$$|GHZ\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}} \quad \text{and} \quad |W\rangle = \frac{1}{\sqrt{3}} (|001\rangle + |010\rangle + |100\rangle).$$

The squashed entanglement of the the general GHZ state is

$$\begin{aligned}
E_{\text{sq}}(X_1; X_2; \dots; X_m)_{GHZ} &= \frac{1}{2} \inf_E I(X_1; X_2; \dots; X_m | E) \\
&= \frac{1}{2} I(X_1; X_2; \dots; X_m) \quad (\text{pure state}) \\
&= \frac{1}{2} \left[\sum_i^m \underbrace{H(X_i)}_{\text{max. mixed}} - \underbrace{H(X_1 \dots X_m)}_{\text{pure}} \right] \\
&= \frac{m}{2}
\end{aligned}$$

For the W state, the 1-qubit reduced systems are of the form

$$\text{Tr}_{X_2 \dots X_m} (|W\rangle\langle W|) = \begin{pmatrix} \frac{m-1}{m} & 0 \\ 0 & \frac{1}{m} \end{pmatrix} \quad (4.38)$$

and so the squashed entanglement for the W state is given by the formula

$$\begin{aligned}
E_{\text{sq}}(X_1; X_2; \dots; X_m)_W &= \frac{1}{2} I(X_1; X_2; \dots; X_m) \\
&= \frac{1}{2} \left[\sum_i^m H(X_i) - \underbrace{H(X_1 \dots X_m)}_{=0} \right] \\
&= \frac{m}{2} \log_2 \left(\frac{m}{(m-1)^{\frac{(m-1)}{m}}} \right) \\
&= \frac{1}{2} \log_2 \left(\frac{m^m}{(m-1)^{(m-1)}} \right) \\
&= \frac{1}{2} \log_2 m + O(1) \ll \frac{m}{2}.
\end{aligned}$$

We can see that the GHZ state is maximally multiparty entangled whereas the W state contains very little multiparty entanglement.

CHAPTER 5

Multiparty distributed compression

Distributed compression of classical information, as discussed in Section 2.1.6, involves many parties collaboratively encoding their classical sources $X_1, X_2 \cdots X_m$ and sending the information to a common receiver [34]. In the quantum setting, the parties are given a quantum state $\varphi^{A_1 A_2 \cdots A_m} \in \mathcal{H}^{A_1 A_2 \cdots A_m}$ and are asked to individually compress their shares of the state and transfer them to the receiver while sending as few qubits as possible [14]. We have already discussed a version of quantum distributed compression in Section 3.3 where we used shared entanglement and classical communication to accomplish the task [43]. In this chapter, we consider the fully quantum scenario where only quantum communication is used and classical communication is forbidden.

In our analysis, we work in the case where we have many copies of the input state, so that the goal is to send shares of the purification $|\psi\rangle^{A_1 A_2 \cdots A_m R} = (|\varphi\rangle^{A_1 A_2 \cdots A_m R})^{\otimes n}$, where the A_i 's denote the m different systems and R denotes the reference system, which does not participate in the protocol. A word on notation is in order. We use A_i to denote both the individual system associated with state φ as well the n -copy version $A_i^{\otimes n}$ associated with ψ ; the intended meaning should be clear from the context. We also use the shorthand notation $A = A_1 A_2 \cdots A_m$ to denote all the senders.

The objective of distributed compression is for the participants to transfer their R -entanglement to a third party Charlie as illustrated in Figure 5–1. As discussed in Section 2.2.5, preserving the R -entanglement means our protocol has high *entanglement fidelity* [41] which guarantees that we can transfer the

state $\varphi^{A_1 A_2 \dots A_m}$, but also preserve all the correlations this state has with the rest of the world.

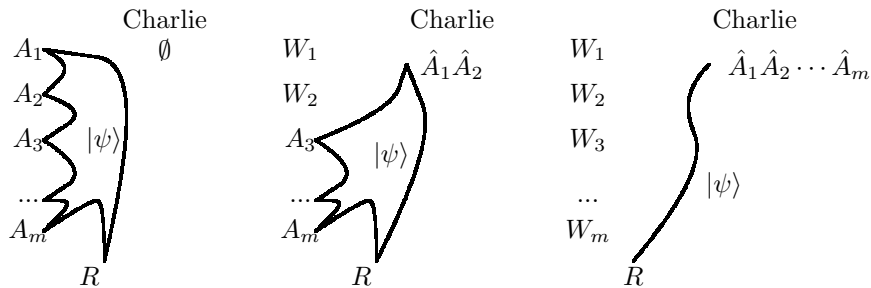


Figure 5–1: Pictorial representation of the quantum correlations between the systems at three stages of the protocol. Originally the state $|\psi\rangle$ is shared between $A_1 A_2 \dots A_m$ and R . The middle picture shows the protocol in progress. Finally, all systems are received by Charlie and $|\psi\rangle$ is now shared between Charlie’s systems $\hat{A}_1 \hat{A}_2 \dots \hat{A}_m$ and R .

An equivalent way of thinking about quantum distributed compression is to say that the participants are attempting to decouple their systems from the reference R solely by sending quantum information to Charlie. Indeed, if we assume that originally R is the purification of $A_1 A_2 \dots A_m$, and at the end of the protocol there are no correlations between the remnant W systems (see Figure 5–1) and R , then the purification of R must have been transferred to Charlie’s laboratory since none of the original information was discarded.

To perform the distributed compression task, each of the senders independently encodes her share before sending part of it to Charlie. The encoding operations are modeled by quantum operations (CPTP maps) \mathcal{E}_i with outputs C_i of dimension 2^{nQ_i} . Once Charlie receives the systems that were sent to him, he will apply a decoding operation \mathcal{D} , with output system $\hat{A} = \hat{A}_1 \hat{A}_2 \dots \hat{A}_m$ isomorphic to the original $A = A_1 A_2 \dots A_m$.

Definition 5.1 (The rate region). *We say that a rate tuple $\vec{Q} = (Q_1, Q_2, \dots, Q_m)$ is achievable if for all $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n \geq N(\epsilon)$ there exist n -dependent maps $(\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_m, \mathcal{D})$ with domains and ranges as*

in the previous paragraph for which the fidelity between the original state, $|\psi\rangle^{A^n R^n} = \left(|\varphi\rangle^{A_1 A_2 \dots A_m R}\right)^{\otimes n}$, and the final state, $\sigma^{\hat{A}_1 \hat{A}_2 \dots \hat{A}_m R} = \sigma^{\hat{A}^n R^n}$, satisfies

$$F\left(|\psi\rangle^{A^n R^n}, \sigma^{\hat{A}^n R^n}\right) = \hat{A}^n R^n \langle \psi | (\mathcal{D} \circ (\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_m)) (\psi^{A^n R^n}) | \psi \rangle^{\hat{A}^n R^n} \geq 1 - \epsilon.$$

We call the closure of the set of achievable rate tuples the rate region.

5.1 The multiparty FQSW protocol

Like the original FQSW protocol, the multiparty version relies on Schumacher compression and the mixing effect of random unitary operations for the encoding. The only additional ingredient is an agreed upon permutation of the participants. The temporal order in which the participants will perform their encoding is of no importance. However, the permutation determines how much information each participant is to send to Charlie.

For each permutation π of the participants, the protocol consists of the following steps:

1. Each Alice- i performs Schumacher compression on her system A_i reducing its effective size to the entropy bound of roughly $H(A_i)$ qubits per copy of the state.
2. Each participant applies a known, pre-selected random unitary to the compressed system.
3. Participant i sends to Charlie a system C_i of dimension 2^{nQ_i} where

$$Q_i > \frac{1}{2} I(A_i; A_{\mathcal{K}_i} R)_\varphi \quad (5.1)$$

where $\mathcal{K}_i = \{\pi(j) : j > \pi^{-1}(i)\}$ is the set of participants who come after participant i according to the permutation.

4. Charlie applies a decoding operation D consisting of the composition of the decoding maps $\mathcal{D}_{\pi(m)} \circ \cdots \circ \mathcal{D}_{\pi(2)} \circ \mathcal{D}_{\pi(1)}$ defined by the individual FQSW steps in order to recover $\sigma^{\hat{A}_1 \hat{A}_2 \cdots \hat{A}_m}$ nearly identical to the original $\psi^{A_1 A_2 \cdots A_m}$ and purifying R .

Note that, in order to perform the decoding operation \mathcal{D} , Charlie needs to know which random unitaries which were used in the individual encoding operations \mathcal{E}_i . We assume this information is shared before the beginning of the protocol in addition to the permutation π .

5.1.1 Statement of results

This section contains our two main theorems about multiparty distributed compression. In Theorem 5.2 we give the formula for the set of achievable rates using the multiparty FQSW protocol (sufficient conditions). Then, in Theorem 5.3 we specify another set of inequalities for the rates Q_i which must be true for any distributed compression protocol (necessary conditions). In what follows, we consistently use $\mathcal{K} \subseteq \{1, 2, \dots, m\}$ to denote any subset of the senders in the protocol.

Theorem 5.2. *Let $|\varphi\rangle^{A_1 A_2 \cdots A_m R}$ be a pure state. If the inequality*

$$\sum_{k \in \mathcal{K}} Q_k \geq \frac{1}{2} \left[\sum_{k \in \mathcal{K}} [H(A_k)_\varphi] + H(R)_\varphi - H(RA_{\mathcal{K}})_\varphi \right] \quad (5.2)$$

holds for all $\mathcal{K} \subseteq \{1, 2, \dots, m\}$, then the rate tuple (Q_1, Q_2, \dots, Q_m) is achievable for distributed compression of the A_i systems.

Because Theorem 5.2 expresses a set of sufficient conditions for the protocol to succeed, we say that these rates are contained in the rate region. The proof is given in the next section.

In the m -dimensional space of rate tuples $(Q_1, Q_2, \dots, Q_m) \in \mathbb{R}^m$, the inequalities (5.2) define a convex polyhedron [69] whose facets are given by the corresponding hyperplanes, as illustrated in Figure 5–2. More specifically,

the rate region is a supermodular polyhedron [70], which means that it has some special properties that will help us in the proof of Theorem 5.2.

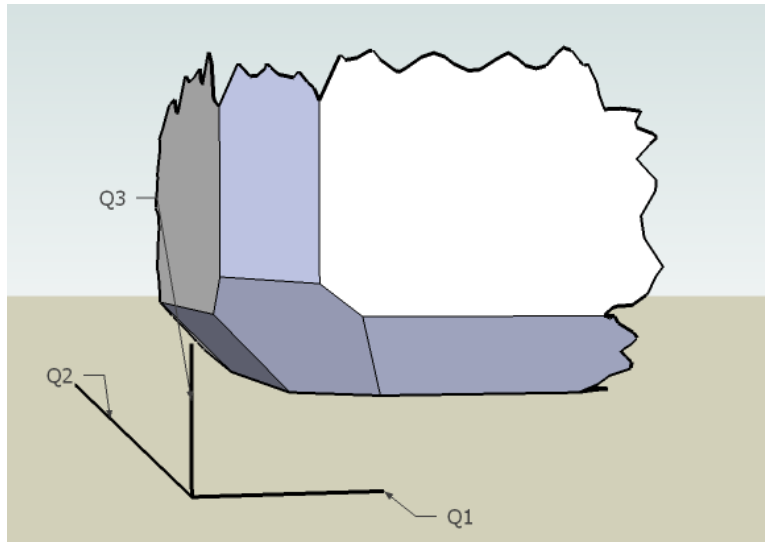


Figure 5-2: The rate region for the multiparty FQSW protocol with three senders.

In order to characterize the rate region further we formulate Theorem 5.3, an outer bound on the rates that must be satisfied for *all* distributed compression protocols.

Theorem 5.3. *Let $|\varphi\rangle^{A_1 A_2 \dots A_m R}$ be a pure state input to a distributed compression protocol which achieves the rate tuple (Q_1, Q_2, \dots, Q_m) , then it must be true that*

$$\sum_{k \in \mathcal{K}} Q_k \geq \frac{1}{2} \left[\sum_{k \in \mathcal{K}} [H(A_k)_\varphi] + H(R)_\varphi - H(RA_{\mathcal{K}})_\varphi \right] - E_{sq}(A_{k_1}; A_{k_2}; \dots; A_{k_{|\mathcal{K}|}})_\varphi, \quad (5.3)$$

for all $\mathcal{K} \subseteq \{1, 2, \dots, m\}$, where E_{sq} is the multiparty squashed entanglement.

The multiparty squashed entanglement was defined in Section 4.2 above.

Notice that Theorems 5.2 and 5.3 both provide bounds of the same form and only differ by the presence of the E_{sq} term. The rate region is squeezed somewhere between these two bounds as illustrated in Figure 5–3.

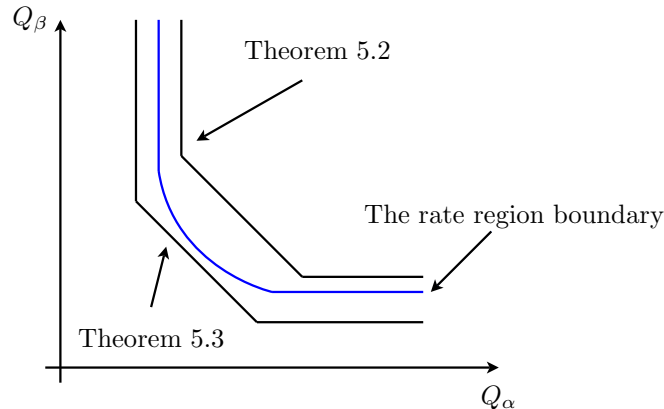


Figure 5–3: A two dimensional diagram showing the inner bound from Theorem 5.2 and the outer bound from Theorem 5.3. The boundary of the real rate region must lie somewhere in between.

For states which have zero squashed entanglement, the inner and outer bounds on the region coincide so that in those cases our protocol is an optimal solution to the multiparty distributed compression problem.

5.2 Proof of inner bound

The multiparty fully quantum Slepian-Wolf protocol can be constructed directly [71] or through the repeated application of the two-party FQSW protocol [5]. We choose the latter approach here in order to illustrate the power of the FQSW protocol as a building block for more complex protocols. To complete the proof we will have to “stitch together” different achievable points using some concepts from the theory of polyhedra [69]. The multiparty rate region has a complex but regular geometry so it is important that we use the right language to describe it. The geometry of multiparty rate regions has previously been discussed in [70, 72].

For every permutation $\pi \in S_m$ of the m senders, there is a different rate tuple $\vec{q}_\pi = (Q_1, Q_2, \dots, Q_m)_\pi \in \mathbb{R}^m$ which is achievable in the limit of many copies of the state. By time-sharing we can achieve any rate that lies in the *convex hull* of these points. We will show that the rate region for an input state $|\varphi\rangle^{A_1 \cdots A_m R}$ can equivalently be described by the set of inequalities from Theorem 5.2, that is

$$\sum_{k \in \mathcal{K}} Q_k \geq \frac{1}{2} \left[\sum_{k \in \mathcal{K}} H(A_k)_\varphi + H(R)_\varphi - H(RA_{\mathcal{K}})_\varphi \right] =: C_{\mathcal{K}} \quad (5.4)$$

where $\mathcal{K} \subseteq \{1, 2, \dots, m\}$ ranges over all subsets of participants and $C_{\mathcal{K}}$ is the name we give to the constant on the right hand side of the inequality. The proof of Theorem 5.2 proceeds in two steps. First we show the set of rate tuples $\{\vec{q}_\pi\}$ is contained in the rate region and then we prove that the set of inequalities (5.4) is an equivalent description of the rates obtained by time sharing and resource wasting of the rates $\{\vec{q}_\pi\}$.

Consider the m -dimensional space of rate tuples $(Q_1, \dots, Q_m) \in \mathbb{R}^m$. We begin by a formal definition of a corner point \vec{q}_π .

Definition 5.4 (Corner point). *Let $\pi \in S_m$ be a permutations of the senders in the protocol. The corresponding rate tuple $q_\pi = (Q_1, Q_2, \dots, Q_m)$ is a corner point if*

$$Q_{\pi(k)} = \frac{1}{2} I(A_{\pi(k)}; A_{\pi(k+1)} \cdots A_{\pi(m)} R) \quad (5.5)$$

where the set $A_{\pi(k+1)} \cdots A_{\pi(m)}$ denotes all the systems which come after k in the permutation π .

We define $\mathcal{Q} := \{\vec{q}_\pi : \pi \in S_m\}$, the set of all corner points. Clearly, $|\mathcal{Q}| \leq m!$ but since some permutations might lead to the same rate tuple, the inequality may be strict.

Lemma 5.5. *The set of corner points, $\mathcal{Q} = \{\vec{q}_\pi : \pi \in S_m\}$, is contained in the rate region.*

Proof sketch for Lemma 5.5. We will now exhibit a protocol that achieves one such point. In order to simplify the notation, but without loss of generality, we choose the reversed-order permutation $\pi = (m, \dots, 2, 1)$. This choice of permutation corresponds to Alice- m sending her information first and Alice-1 sending last.

We will repeatedly use the FQSW protocol in order to send the m systems to Charlie:

1. The first party Schumacher compresses her system A_m and sends it to Charlie. She succeeds provided

$$Q_m \geq \frac{1}{2}I(A_m; A_1 A_2 \dots A_{m-1} R) + \delta = H(A_m) + \delta$$

for any $\delta > 0$. The above rate is dictated by the FQSW inequality (3.25) because we are facing the same type of problem except that the “reference” consists of R as well as the remaining participants $A_1 A_2 \dots A_{m-1}$. The fact that the formula reduces to $Q_m > H(A_m)$ should also be expected since there are no correlations that the first participant can take advantage of; she is just performing Schumacher compression.

2. The second party also faces an instance of an FQSW problem. The task is to transmit the system A_{m-1} to Charlie, who is now assumed to hold A_m . The purifying system consists of $A_1 A_2 \dots A_{m-2} R$. According to inequality (3.25) the rate must be

$$Q_{m-1} \geq \frac{1}{2}I(A_{m-1}; A_1 A_2 \dots A_{m-2} R) + \delta$$

for any $\delta > 0$.

3. The last person to be merging with Charlie will have a purifying system consisting of only R . Her transfer will be successful if

$$Q_1 \geq \frac{1}{2}I(A_1; R) + \delta$$

for any $\delta > 0$.

On the receiving end of the protocol, Charlie will apply the decoding map \mathcal{D} consisting of the composition of the decoding maps $\mathcal{D}_1 \circ \mathcal{D}_2 \circ \dots \circ \mathcal{D}_m$ defined by the individual FQSW steps to recover the state $\sigma^{\hat{A}_1 \hat{A}_2 \dots \hat{A}_m}$, which will be such that the fidelity between $|\psi\rangle^{A^n R^n}$ and $\sigma^{\hat{A}^n R^n}$ is high, essentially by the triangle inequality. Finally, because we can make δ arbitrarily small, the rate tuple (Q_1, \dots, Q_m) , with

$$Q_k = \frac{1}{2}I(A_k; A_1 \dots A_{k-1} R), \quad (5.6)$$

must be contained in the rate region. The same argument applies for each permutation $\pi \in S_m$, leading to the conclusion that the full set \mathcal{Q} is contained in the rate region. \square

Each one of the corner points \vec{q}_π can also be described by an equivalent set of equations involving sums of the rates.

Lemma 5.6. *The rate tuple (Q_1, Q_2, \dots, Q_m) is a corner point if and only if for some $\pi \in S_m$ and for all l such that $1 \leq l \leq m$,*

$$\sum_{m-l+1 \leq k \leq m} Q_{\pi(k)} = \frac{1}{2} \left[\sum_{m-l+1 \leq k \leq m} H(A_{\pi(k)}) + H(R) - H(A_{\pi[m-l+1, m]} R) \right] = C_{\pi[m-l+1, m]} \quad (5.7)$$

where $A_{\pi[m-l+1, m]} := A_{\pi(m-l+1)} A_{\pi(m-l+2)} \dots A_{\pi(m)}$ denotes the last l participants according to the permutation π .

Proof of Lemma 5.6. The proof follows trivially from Lemma 5.5 by considering sums of the rates. If we again choose the permutation $\pi = (m, \dots, 2, 1)$

for simplicity, we see that the sum of the rates of the last l participants is

$$\begin{aligned} Q_1 + \cdots + Q_l &= \frac{1}{2} \left[I(A_1; R) + I(A_2; A_1 R) + \cdots + I(A_l; A_1 \cdots A_{l-1} R) \right] \\ &= \frac{1}{2} \left[\sum_{1 \leq k \leq l} H(A_k) + H(R) - H(A_1 \cdots A_l R) \right] = C_{12 \dots l}. \end{aligned} \quad (5.8)$$

A telescoping effect occurs and most of the inner terms cancel so we are left with a system of equations identical to (5.7). Moreover, this system is clearly solvable for the individual rates Q_k . The analogous simplification occurs for all other permutations. \square

So far, we have shown that the set of corner points \mathcal{Q} is contained in the rate region of the multiparty fully quantum Slepian-Wolf protocol. The convex hull of a set of points \mathcal{Q} is defined to be

$$\text{conv}(\mathcal{Q}) := \left\{ \vec{x} \in \mathbb{R}^m : \vec{x} = \sum \lambda_i \vec{q}_i, \vec{q}_i \in \mathcal{Q}, \lambda_i \geq 0, \sum \lambda_i = 1 \right\}. \quad (5.9)$$

Because of the possibility of time-sharing between the different corner points, the entire convex hull $\text{conv}(\mathcal{Q})$ must be achievable. Furthermore, by simply allowing any one of the senders to waste resources, we know that if a rate tuple \vec{q} is achievable, then so is $\vec{q} + \vec{w}$ for any vector \vec{w} with nonnegative coefficients. More formally, we say that any $\vec{q} + \text{cone}(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m)$ is also inside the rate region, where $\{\vec{e}_i\}$ is the standard basis for \mathbb{R}^m : $\vec{e}_i = (0, 0, \dots, \underbrace{0, 1, 0, 0}_i, 0, 0)$ and

$$\text{cone}(\vec{e}_1, \dots, \vec{e}_m) := \left\{ \vec{x} \in \mathbb{R}^m : \vec{x} = \sum \lambda_i \vec{e}_i, \lambda_i \geq 0 \right\}. \quad (5.10)$$

Thus, we have demonstrated that the set of rates

$$P_{\mathcal{V}} := \text{conv}(\mathcal{Q}) + \text{cone}(\vec{e}_1, \dots, \vec{e}_m) \quad (5.11)$$

is achievable. To complete the proof of Theorem 5.2, we will need to show that $P_{\mathcal{V}}$ has an equivalent description as

$$P_{\mathcal{H}} := \left\{ (Q_1, \dots, Q_m) \in \mathbb{R}^m : \sum_{k \in \mathcal{K}} Q_k \geq C_{\mathcal{K}}, \forall \mathcal{K} \subseteq \{1, 2, \dots, m\} \right\}, \quad (5.12)$$

where the constants $C_{\mathcal{K}}$ are as defined in equation (5.4). This equivalence is an explicit special case of the Minkowski-Weyl Theorem on convex polyhedra.

Theorem 5.7 (Minkowski-Weyl Theorem). *[69, p.30] For a subset $P \subseteq \mathbb{R}^m$, the following two statements are equivalent:*

- *P is a \mathcal{V} -polyhedron: the sum of a convex hull of a finite set of points $\mathcal{P} = \{\vec{p}_i\}$ plus a conical combination of vectors $\mathcal{W} = \{\vec{w}_i\}$*

$$P = \text{conv}(\mathcal{P}) + \text{cone}(\mathcal{W}) \quad (5.13)$$

where $\text{conv}(\mathcal{P})$ and $\text{cone}(\mathcal{W})$ are defined in (5.9) and (5.10) respectively.

- *P is a \mathcal{H} -polyhedron: an intersection of n closed halfspaces*

$$P = \{\vec{x} \in \mathbb{R}^m : A\vec{x} \geq \vec{a}\} \quad (5.14)$$

for some matrix $A \in \mathbb{R}^{n \times m}$ and some vector $\vec{a} \in \mathbb{R}^n$. Each of the n rows in equation (5.14) defines one halfspace.

Preliminaries Before we begin the equivalence proof in earnest, we make two useful observations which will be instrumental to our subsequent argument. First, we prove a very important property of the constants $C_{\mathcal{K}}$ which will dictate the geometry of the rate region.

Lemma 5.8 (Superadditivity). *Let $\mathcal{K}, \mathcal{L} \subseteq \{1, 2, \dots, m\}$ be any two subsets of the senders. Then*

$$C_{\mathcal{K} \cup \mathcal{L}} + C_{\mathcal{K} \cap \mathcal{L}} \geq C_{\mathcal{K}} + C_{\mathcal{L}}. \quad (5.15)$$

Proof of Lemma 5.8. We expand the C terms and cancel the $\frac{1}{2}$ -factors to obtain

$$\begin{aligned} \sum_{k \in \mathcal{K} \cup \mathcal{L}} H(A_k) + H(R) - H(RA_{\mathcal{K} \cup \mathcal{L}}) &\geq \sum_{k \in \mathcal{K}} H(A_k) + H(R) - H(RA_{\mathcal{K}}) \\ &+ \sum_{k \in \mathcal{K} \cap \mathcal{L}} H(A_k) + H(R) - H(RA_{\mathcal{K} \cap \mathcal{L}}) \geq \sum_{k \in \mathcal{L}} H(A_k) + H(R) - H(RA_{\mathcal{L}}). \end{aligned}$$

After canceling all common terms we find that the above inequality is equivalent to

$$H(RA_{\mathcal{K}}) + H(RA_{\mathcal{L}}) \geq H(RA_{\mathcal{K} \cup \mathcal{L}}) + H(RA_{\mathcal{K} \cap \mathcal{L}}), \quad (5.16)$$

which is true by the strong subadditivity (SSA) inequality of quantum entropy [36]. \square

As a consequence of this lemma, we can derive an equivalence property for the saturated inequalities.

Corollary 5.9. *Suppose that the following two equations hold for a given point of $P_{\mathcal{H}}$:*

$$\sum_{k \in \mathcal{K}} Q_k = C_{\mathcal{K}} \quad \text{and} \quad \sum_{k \in \mathcal{L}} Q_k = C_{\mathcal{L}}. \quad (5.17)$$

Then the following equations must also be true:

$$\sum_{k \in \mathcal{K} \cup \mathcal{L}} Q_k = C_{\mathcal{K} \cup \mathcal{L}} \quad \text{and} \quad \sum_{k \in \mathcal{K} \cap \mathcal{L}} Q_k = C_{\mathcal{K} \cap \mathcal{L}}. \quad (5.18)$$

Proof of Corollary 5.9. The proof follows from the equation

$$\sum_{k \in \mathcal{K}} Q_k + \sum_{k \in \mathcal{L}} Q_k = C_{\mathcal{K}} + C_{\mathcal{L}} \leq C_{\mathcal{K} \cup \mathcal{L}} + C_{\mathcal{K} \cap \mathcal{L}} \leq \sum_{k \in \mathcal{K} \cup \mathcal{L}} Q_k + \sum_{k \in \mathcal{K} \cap \mathcal{L}} Q_k \quad (5.19)$$

where the first inequality comes from Lemma 5.8. The second inequality is true by the definition of $P_{\mathcal{H}}$ since $\mathcal{K} \cup \mathcal{L}$ and $\mathcal{K} \cap \mathcal{L}$ are subsets of $\{1, 2, \dots, m\}$. Because the leftmost terms and rightmost terms are identical, we must have

equality throughout equation (5.19), which in turn implies the the union and the intersection equations are saturated. \square

An important consequence of Lemma 5.8 is that it implies that the polyhedron $P_{\mathcal{H}}$ has a very special structure. It is known as a supermodular polyhedron or contra-polymatroid. The fact that $\text{conv}(Q) = P_{\mathcal{H}}$ was proved by Edmonds [70], whose ingenious proof makes use of linear programming duality. Below we give an elementary proof that does not use duality.

A *vertex* is a zero-dimensional face of a polyhedron. A point $\bar{Q} = (\bar{Q}_1, \bar{Q}_2, \dots, \bar{Q}_m) \in P_{\mathcal{H}} \subset \mathbb{R}^m$ is a vertex of $P_{\mathcal{H}}$ if and only if it is the unique solution of a set of linearly independent equations

$$\sum_{k \in \mathcal{L}_i} Q_k = C_{\mathcal{L}_i}, \quad 1 \leq i \leq m \quad (5.20)$$

for some subsets $\mathcal{L}_i \subseteq \{1, 2, \dots, m\}$. In the remainder of the proof we require only a specific consequence of linear independence, which we state in the following lemma.

Lemma 5.10 (No co-occurrence). *Let $\mathcal{L}_i \subseteq \{1, 2, \dots, m\}$ be a collection of m sets such that the system (5.20) has a unique solution. Then there is no pair of elements j, k such that $j \in \mathcal{L}_i$ if and only if $k \in \mathcal{L}_i$ for all i .*

Proof. If there was such a pair j and k , then the corresponding columns of the left hand side of (5.20) would be linearly dependent. \square

Armed with the above tools, we will now show that there is a one-to-one correspondence between the corner points Q and the vertices of the \mathcal{H} -polyhedron $P_{\mathcal{H}}$. We will then show that the vectors that generate the cone part of the \mathcal{H} -polyhedron correspond to the resource wasting vectors $\{\vec{e}_i\}$.

Step 1: $\mathcal{Q} \subseteq \text{vertices}(P_{\mathcal{H}})$ We know from Lemma 5.6 that every point $\vec{q}_{\pi} \in \mathcal{Q}$ satisfies the m equations

$$\sum_{m-i+1 \leq k \leq m} Q_{\pi(k)} = C_{\pi[m-i+1, m]}, \quad 1 \leq i \leq m. \quad (5.21)$$

The equations (5.21) are linearly independent since the left hand side is triangular, and have the form of the inequalities in (5.12) that are used to define $P_{\mathcal{H}}$. They have the unique solution:

$$Q_{\pi(m)} = C_{\pi(m)} \quad Q_{\pi(i)} = C_{\pi[i, m]} - C_{\pi[i+1, m]}, \quad 1 \leq i \leq m-1. \quad (5.22)$$

We need to show that this solution satisfies all the inequalities used to define $P_{\mathcal{H}}$ in (5.12). We proceed by induction on $|\mathcal{K}|$. The case $|\mathcal{K}| = 1$ follows from (5.22) and the superadditivity property (5.15). For $|\mathcal{K}| \geq 2$ we can write $\mathcal{K} = \{\pi(i)\} \cup \mathcal{K}'$ for some $\mathcal{K}' \subseteq \{\pi(i+1), \pi(i+2), \dots, \pi(m)\}$. Then

$$\begin{aligned} \sum_{k \in \mathcal{K}} Q_k &= Q_{\pi(i)} + \sum_{k \in \mathcal{K}'} Q_k \\ &\geq C_{\pi[i, m]} - C_{\pi[i+1, m]} + \sum_{k \in \mathcal{K}'} Q_k \\ &\geq C_{\pi[i, m]} - C_{\pi[i+1, m]} + C_{\mathcal{K}'} \quad (\text{induction}) \\ &\geq C_{\mathcal{K}} \end{aligned}$$

where we again used superadditivity to get the last inequality.

Step 2: $\text{vertices}(P_{\mathcal{H}}) \subseteq \mathcal{Q}$ In order to prove the opposite inclusion, we will show that every vertex of $P_{\mathcal{H}}$ is of the form of Lemma 5.6. More specifically, we want to prove the following proposition.

Proposition 5.11 (Existence of a maximal chain). *Every vertex of $P_{\mathcal{H}}$, that is, the intersection of m linearly independent hyperplanes*

$$\sum_{k \in \mathcal{L}_i} Q_k = C_{\mathcal{L}_i}, \quad 1 \leq i \leq m, \quad (5.23)$$

defined by the family of sets $\{\mathcal{L}_i; 1 \leq i \leq m\}$ can be described by an equivalent set of equations

$$\sum_{k \in \mathcal{K}_i} Q_k = C_{\mathcal{K}_i}, \quad 1 \leq i \leq m, \quad (5.24)$$

for some family of sets distinct $\mathcal{K}_i \subseteq \{1, 2, \dots, m\}$ that form a maximal chain in the sense of

$$\emptyset = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \mathcal{K}_2 \subset \dots \subset \mathcal{K}_{m-1} \subset \mathcal{K}_m = \{1, 2, \dots, m\}. \quad (5.25)$$

Since there exists a permutation π such that $\forall i, \pi[m - i + 1, m] = \mathcal{K}_i$ this implies that all the vertices of $P_{\mathcal{H}}$ are in \mathcal{Q} . The main tool we have at our disposal in order to prove this proposition is Corollary 5.9, which we will use extensively.

Proof of Proposition 5.11. Let $\{\mathcal{L}_i\}_{i=1}^m$ be the subsets of $\{1, 2, \dots, m\}$ for which the inequalities are saturated and define $\mathcal{L}_i^{\mathcal{S}} := \mathcal{L}_i \cap \mathcal{S}$, the intersection of \mathcal{L}_i with some set $\mathcal{S} \subseteq \{1, 2, \dots, m\}$.

Construct the directed graph $G = (V, E)$, where:

- $V = \{1, 2, \dots, m\}$, i.e. the vertices are the numbers from 1 to m ;
- $E = \{(j, k) : (\forall i) j \in \mathcal{L}_i \implies k \in \mathcal{L}_i\}$, i.e. there is an edge from vertex j to vertex k if whenever vertex j occurs in the given subsets, then so does vertex k .

Now G has to be acyclic by Lemma 5.10, so it has a topological sorted order.

Let us call this order ν . Let $\mathcal{K}_0 = \emptyset$ and let

$$\mathcal{K}_l = \{\nu_{m-l+1}, \dots, \nu_m\} \quad (5.26)$$

for $l \in \{1, \dots, m\}$. The sets \mathcal{K}_l , which consist of the last l vertices according to the ordering ν , form a maximal chain $\mathcal{K}_0 \subset \mathcal{K}_1 \subset \dots \subset \mathcal{K}_{m-1} \subset \mathcal{K}_m$ by construction.

We claim that all the sets \mathcal{K}_l can be constructed from the sets $\{\mathcal{L}_i\}$ by using unions and intersections as dictated by Corollary 5.9. The statement is true for $\mathcal{K}_m = \{1, 2, \dots, m\}$ because every variable must appear in some constraint equation, giving $\mathcal{K}_m = \cup_i \mathcal{L}_i$. The statement is also true for $\mathcal{K}_{m-1} = \{\nu_2, \dots, \nu_m\}$ since the vertex ν_1 has no in-edges in G by the definition of a topological sort, which means that

$$\mathcal{K}_{m-1} = \bigcup_{\nu_1 \notin \mathcal{L}_i^{\mathcal{K}_m}} \mathcal{L}_i^{\mathcal{K}_m}. \quad (5.27)$$

For the induction statement, let $l \in \{m-1, \dots, 2, 1\}$ and assume that $\mathcal{K}_l = \bigcup_i \mathcal{L}_i^{\mathcal{K}_l}$. Since the vertex ν_{m-l} has no in-edges in the induced subgraph generated by the vertices \mathcal{K}_l by the definition of the topological sort, \mathcal{K}_{l-1} can be obtained from the union of all the sets not containing ν_{m-l} :

$$\mathcal{K}_{l-1} = \bigcup_{\nu_{m-l} \notin \mathcal{L}_i^{\mathcal{K}_l}} \mathcal{L}_i^{\mathcal{K}_l}. \quad (5.28)$$

In more detail, we claim that for all $\omega \neq \nu_{m-l} \in \mathcal{K}_{l-1}$ there exists i such that $\nu_{m-l} \notin \mathcal{L}_i^{\mathcal{K}_l}$ and $\omega \in \mathcal{L}_i^{\mathcal{K}_l}$. If it were not true, that would imply the existence of $\omega \neq \nu_{m-l} \in \mathcal{K}_{l-1}$ such that for all i , $\nu_{m-l} \in \mathcal{L}_i^{\mathcal{K}_l}$ or $\omega \notin \mathcal{L}_i^{\mathcal{K}_l}$. This last condition implies that whenever $\omega \in \mathcal{L}_i^{\mathcal{K}_l}$ it is also true that $\nu_{m-l} \in \mathcal{L}_i^{\mathcal{K}_l}$, which corresponds to an edge (ω, ν_{m-l}) in the induced subgraph. \square

We have shown that every vertex can be written in precisely the same form as Lemma 5.6 and is therefore a point in \mathcal{Q} . This proves $vertices(P_{\mathcal{H}}) \subseteq \mathcal{Q}$, which together with the result of Step 1, implies $vertices(P_{\mathcal{H}}) = \mathcal{Q}$.

Step 3: Cone Part The final step is to find the set of direction vectors that correspond to the cone part of $P_{\mathcal{H}}$. The generating vectors of the cone are all vectors that satisfy the homogeneous versions of the halfspace inequalities

(5.14), which in our case gives

$$\sum_{k \in \mathcal{K}} Q_k \geq 0 \quad (5.29)$$

for all $\mathcal{K} \subset \{1, 2, \dots, m\}$. These inequalities are satisfied if and only if $Q_k \geq 0$ for all k . We can therefore conclude that the cone part of $P_{\mathcal{H}}$ is $\text{cone}(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m)$.

This completes our demonstration that $P_{\mathcal{V}}$ is the \mathcal{V} -polyhedron description of the \mathcal{H} -polyhedron $P_{\mathcal{H}}$. Thus we arrive at the statement we were trying to prove; if the inequalities

$$\sum_{k \in \mathcal{K}} Q_k \geq C_{\mathcal{K}} = \frac{1}{2} \left[\sum_{k \in \mathcal{K}} H(A_k)_{\varphi} + H(R)_{\varphi} - H(RA_{\mathcal{K}})_{\varphi} \right] \quad (5.30)$$

are satisfied for any $\mathcal{K} \subseteq \{1, 2, \dots, m\}$, then the rate tuple (Q_1, Q_2, \dots, Q_m) is inside the rate region. This completes the proof of Theorem 5.2.

5.3 Proof of outer bound

We want to show that *any* distributed compression protocol which works must satisfy all of the inequalities (5.3) from Theorem 5.3. In order to prove this, we will use some of the properties of multiparty information and squashed entanglement. We break up the proof into three steps.

Step 1: Decoupling Formula We know that the input system $|\psi\rangle^{A^n R^n}$ is a pure state. If we account for the Stinespring dilations of each encoding and decoding operation, then we can view any protocol as implemented by unitary transformations with ancilla and waste. Therefore, the output state (including the waste systems) should also be pure. More specifically, the encoding operations are modeled by CPTP maps \mathcal{E}_i with outputs C_i of dimension 2^{nQ_i} . In our analysis we will keep track of the purification (waste) systems W_i of the

the Stinespring dilations \mathcal{E}_i , so the evolution as a whole will be unitary.

$$\begin{array}{c} A_i \text{ --- } \boxed{\mathcal{E}_i} \text{ --- } C_i \quad \leftarrow \text{to Charlie} \\ |0\rangle \text{ --- } \boxed{\mathcal{E}_i} \text{ --- } W_i \quad \leftarrow \text{waste} \end{array}$$

Once Charlie receives the systems that were sent to him, he will apply a decoding CPTP map \mathcal{D} with output system $\hat{A} = \hat{A}_1 \hat{A}_2 \dots \hat{A}_m$ isomorphic to the original $A = A_1 A_2 \dots A_m$.

$$\begin{array}{c} \bigcup_i^m C_i \text{ --- } \boxed{\mathcal{D}} \text{ --- } \hat{A}_1 \dots \hat{A}_m \quad \leftarrow \text{near-purification of } R \\ |0\rangle \text{ --- } \boxed{\mathcal{D}} \text{ --- } W_C \quad \leftarrow \text{Charlie's waste} \end{array}$$

In what follows we will use Figure 5–4 extensively in order to keep track of the evolution and purity of the states at various points in the protocol.

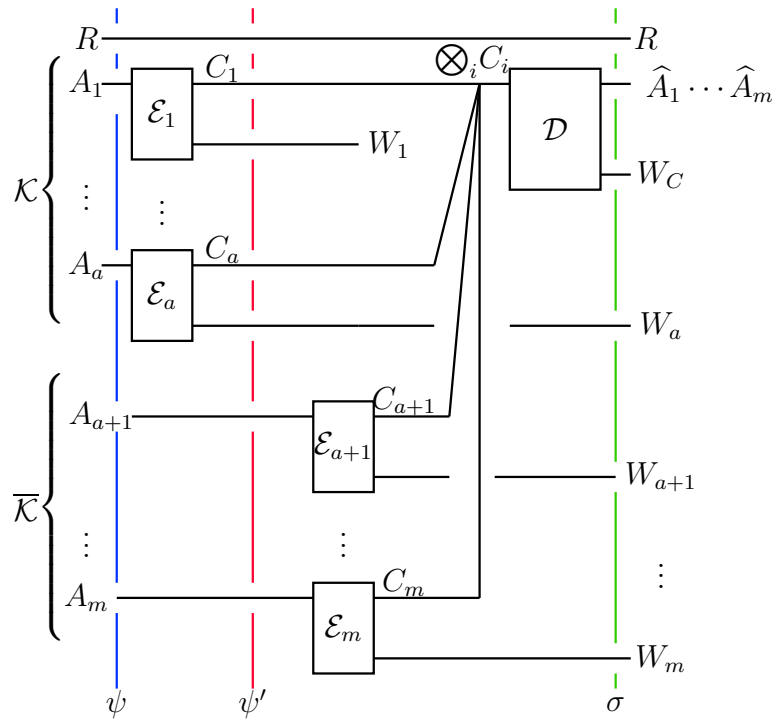


Figure 5–4: A general distributed compression circuit diagram showing the encoding operations \mathcal{E}_i with output systems C_i (compressed data) and W_i (waste). The decoding operation takes all the compressed data $\bigotimes_i C_i$ and applies the decoding operation \mathcal{D} to output a state $\sigma^{\hat{A}^n R^n}$ which has high fidelity with the original $|\psi\rangle^{A^n R^n}$.

The starting point of our argument is the fidelity condition (Definition 5.1) for successful distributed compression, which we restate below for convenience

$$F\left(|\psi\rangle^{A^n R^n}, \sigma^{\hat{A}^n R^n}\right) \geq 1 - \epsilon \quad (5.31)$$

where $|\psi\rangle^{A^n R^n} = \left(|\varphi\rangle^{A_1 A_2 \dots A_m R}\right)^{\otimes n}$ is the input state to the protocol and $\sigma^{\hat{A}^n R^n}$ is the output state of the protocol. Since $\sigma^{\hat{A}^n R^n}$ has high fidelity with a rank one state, it must have one large eigenvalue

$$\lambda_{\max}(\sigma^{\hat{A}^n R^n}) \geq 1 - \epsilon. \quad (5.32)$$

Therefore, the full output state $|\sigma\rangle^{\hat{A}^n R^n W_1 \dots W_m W_C}$ has Schmidt decomposition of the form

$$|\sigma\rangle^{\hat{A}^n R^n W_1 \dots W_m W_C} = \sum_i \sqrt{\lambda_i} |e_i\rangle^{\hat{A}^n R^n} \otimes |f_i\rangle^{W_1 \dots W_m W_C}, \quad (5.33)$$

where $|e_i\rangle, |f_i\rangle$ are orthonormal bases and $\lambda_1 = \lambda_{\max} \geq 1 - \epsilon$.

Next we show that the output state $|\sigma\rangle^{\hat{A}^n R^n W_1 \dots W_m W_C}$ is very close in fidelity to a totally decoupled state $\sigma^{\hat{A}^n R^n} \otimes \sigma^{W_1 \dots W_m W_C}$, which is a tensor product of the marginals of $|\sigma\rangle$ on the subsystems $\hat{A}^n R^n$ and $W_1 \dots W_m W_C$:

$$\begin{aligned} F(|\sigma\rangle^{\hat{A}^n R^n W_1 \dots W_m W_C}, \sigma^{\hat{A}^n R^n} \otimes \sigma^{W_1 \dots W_m W_C}) &= \\ &= \text{Tr} \left[|\sigma\rangle\langle\sigma|^{\hat{A}^n R^n W_1 \dots W_m W_C} \left(\sigma^{\hat{A}^n R^n} \otimes \sigma^{W_1 \dots W_m W_C} \right) \right] \\ &= \sum_i \lambda_i^3 \geq (1 - \epsilon)^3 \geq 1 - 3\epsilon. \end{aligned} \quad (5.34)$$

Using the relationship between fidelity and trace distance [40], we can transform (5.34) into the trace distance bound

$$\left\| |\sigma\rangle\langle\sigma|^{\hat{A}^n R^n W_1 \dots W_m W_C} - \sigma^{\hat{A}^n R^n} \otimes \sigma^{W_1 \dots W_m W_C} \right\|_1 \leq 2\sqrt{3\epsilon}. \quad (5.35)$$

By the contractivity of trace distance, the same equation must be true for any subset of the systems. This bound combined with the Fannes inequality

implies that the entropies taken with respect to the output state are nearly additive:

$$\begin{aligned}
|H(R^n W_{\mathcal{K}})_{\sigma} - H(R^n)_{\sigma} + H(W_{\mathcal{K}})_{\sigma}| &\leq 2\sqrt{3}\epsilon \log(d_{R^n} d_{W_{\mathcal{K}}}) + \eta(2\sqrt{3}\epsilon) \\
&\leq 2\sqrt{3}\epsilon \log(d_{A^n} d_{A_{\mathcal{K}}^{2n}}) + \eta(2\sqrt{3}\epsilon) \\
&\leq 2\sqrt{3}\epsilon n \log(d_A^3) + \eta(2\sqrt{3}\epsilon) \\
&=: f_1(\epsilon, n).
\end{aligned} \tag{5.36}$$

for any subset $\mathcal{K} \subseteq \{1, 2 \dots m\}$ with $\epsilon \leq \frac{1}{12e^2}$ and $\eta(x) = -x \log x$. In the second line we have used the fact that $d_A = d_R$ and exploited the fact that $d_{W_{\mathcal{K}}}$ can be taken less than or equal to $d_{A_{\mathcal{K}}^{2n}}$, the maximum size of an environment required for a quantum operation with inputs and outputs of dimension no larger than $d_{A_{\mathcal{K}}^n}$.

Step 2: Dimension Counting The entropy of any system is bounded above by the logarithm of its dimension. In the case of the systems that participants send to Charlie, this implies that

$$n \sum_{k \in \mathcal{K}} Q_k \geq H(C_{\mathcal{K}})_{\psi'}. \tag{5.37}$$

We can use this fact and the diagram of Figure 5–4 to bound the rates Q_i . First we add $H(A_{\bar{\mathcal{K}}})_{\psi} = H(A_{\bar{\mathcal{K}}})_{\psi'}$ to both sides of equation (5.37) and obtain the inequality

$$H(A_{\bar{\mathcal{K}}})_{\psi} + n \sum_{k \in \mathcal{K}} Q_k \geq H(C_{\mathcal{K}})_{\psi'} + H(A_{\bar{\mathcal{K}}})_{\psi'} \geq H(C_{\mathcal{K}} A_{\bar{\mathcal{K}}})_{\psi'}. \tag{5.38}$$

For each encoding operation, the input system A_i is unitarily related to the outputs $C_i W_i$ so we can write

$$H(A_i)_{\psi} = H(W_i C_i)_{\psi'} \leq H(W_i)_{\psi'} + H(C_i)_{\psi'} \leq H(W_i)_{\psi'} + n Q_i, \tag{5.39}$$

where in the last inequality we have used the dimension bound $H(C_i) \leq nQ_i$. If we collect all the Q_i terms from equations (5.38) and (5.39), we obtain the inequalities

$$n \sum_{i \in \mathcal{K}} Q_i \geq H(C_{\mathcal{K}} A_{\bar{\mathcal{K}}})_{\psi'} - H(A_{\bar{\mathcal{K}}})_{\psi} \quad (5.40)$$

$$n \sum_{i \in \mathcal{K}} Q_i \geq \sum_{i \in \mathcal{K}} H(A_i)_{\psi} - \sum_{i \in \mathcal{K}} H(W_i)_{\psi'}. \quad (5.41)$$

Now add equations (5.40) and (5.41) to get

$$\begin{aligned} 2n \sum_{i \in \mathcal{K}} Q_i &\geq \sum_{i \in \mathcal{K}} H(A_i)_{\psi} - \sum_{i \in \mathcal{K}} H(W_i)_{\psi'} + H(C_{\mathcal{K}} A_{\bar{\mathcal{K}}})_{\psi'} - H(A_{\bar{\mathcal{K}}})_{\psi} \\ &\stackrel{(1)}{=} \sum_{i \in \mathcal{K}} H(A_i)_{\psi} - \sum_{i \in \mathcal{K}} H(W_i)_{\psi'} + H(W_{\mathcal{K}} R^n)_{\psi'} - H(R^n A_{\mathcal{K}})_{\psi} \\ &\stackrel{(2)}{\geq} \sum_{i \in \mathcal{K}} H(A_i)_{\psi} - \sum_{i \in \mathcal{K}} H(W_i)_{\psi'} + H(W_{\mathcal{K}})_{\psi'} + H(R^n)_{\psi'} \\ &\quad - H(R^n A_{\mathcal{K}})_{\psi} - f_1(\epsilon, n) \\ &= \left[\sum_{i \in \mathcal{K}} H(A_i) + H(R^n) - H(R^n A_{\mathcal{K}}) \right]_{\psi} + H(W_{\mathcal{K}})_{\psi'} \\ &\quad - \sum_{i \in \mathcal{K}} H(W_i)_{\psi'} - f_1(\epsilon, n), \end{aligned} \quad (5.42)$$

where the equality ⁽¹⁾ comes about because the two systems $|\psi\rangle^{A_{\mathcal{K}} A_{\bar{\mathcal{K}}} R^n}$ and $|\psi'\rangle^{C_{\mathcal{K}} W_{\mathcal{K}} A_{\bar{\mathcal{K}}} R^n}$ are pure. The inequality (5.36) from Step 1 was used in ⁽²⁾.

Step 3: Squashed Entanglement We would like to have a bound on the extra terms in equation (5.42) that does not depend on the encoding and decoding maps. We can accomplish this if we bound the waste terms $\sum_{i \in \mathcal{K}} H(W_i)_{\sigma} - H(W_{\mathcal{K}})_{\sigma}$ by the squashed entanglement $2E_{\text{sq}}(A_{k_1}; \dots; A_{k_l})_{\psi}$ of the input state for each $\mathcal{K} = \{k_1, k_2, \dots, k_l\} \subseteq \{1, \dots, m\}$ plus some small corrections. The proof requires a continuity statement analogous to (5.36), namely that

$$|H(W_i) - H(W_i|R)| \leq f_2(\epsilon, n) \quad (5.43)$$

where f_2 is some function such that $f_2(\epsilon, n)/n \rightarrow 0$ as $\epsilon \rightarrow 0$. The proof is very similar to that of (5.36) so we omit it.

Furthermore, if we allow an arbitrary transformation $\mathcal{N}^{R \rightarrow E}$ to be applied to the R system, we will obtain some general extension but the analog of equation (5.43) will remain true by the contractivity of the trace distance under CPTP maps. We can therefore write:

$$\begin{aligned}
& \sum_{i \in \mathcal{K}} H(W_i)_\psi - H(W_{\mathcal{K}})_\psi \\
& \leq \sum_{i \in \mathcal{K}} H(W_i|E) - H(W_{\mathcal{K}}|E) + [|\mathcal{K}| + 1]f_2(\epsilon, n) \\
& = I(W_{k_1}; W_{k_2}; \dots; W_{k_l}; E) - I(W_{k_1}; E) - \sum_{i \in \{\mathcal{K} \setminus k_1\}} I(W_i; E) + f'_2(\epsilon, n) \\
& =^{(1)} I(W_{k_1}E; W_{k_2}; \dots; W_{k_l}) - \sum_{i \in \{\mathcal{K} \setminus k_1\}} I(W_i; E) + f'_2(\epsilon, n) \\
& \leq^{(2)} I(A_{k_1}E; W_{k_2}; \dots; W_{k_l}) - \sum_{i \in \{\mathcal{K} \setminus k_1\}} I(W_i; E) + f'_2(\epsilon, n) \\
& =^{(1)} I(A_{k_1}; W_{k_2}; \dots; W_{k_l}, E) - I(A_{k_1}; E) - \sum_{i \in \{\mathcal{K} \setminus k_1\}} I(W_i; E) + f'_2(\epsilon, n) \\
& \leq^{(3)} I(A_{k_1}; A_{k_2}; \dots; A_{k_l}; E) - \sum_{i \in \mathcal{K}} I(A_i; E) + f'_2(\epsilon, n) \\
& \leq I(A_{k_1}; A_{k_2}; \dots; A_{k_l}|E) + f'_2(\epsilon, n),
\end{aligned}$$

where we have used the shorthand $f'_2(\epsilon, n) = [|\mathcal{K}| + 1]f_2(\epsilon, n)$ for brevity. Equations marked ⁽¹⁾ use Lemma 4.2 and inequality ⁽²⁾ comes about from Lemma 4.3, the monotonicity of the multiparty information. Inequality ⁽³⁾ is obtained when we repeat the steps for k_2, \dots, k_l . The above result is true for any extension E but we want to find the tightest possible lower bound for the rate region so we take the infimum over all possible extensions E thus arriving at the definition of squashed entanglement.

Putting together equation (5.42) from Step 2 and the bound from Step 3 we have

$$\begin{aligned}
2n \sum_{i \in \mathcal{K}} Q_i &\geq \left[\sum_{i \in \mathcal{K}} H(A_i) + H(R^n) - H(R^n A_{\mathcal{K}}) \right] \\
&\quad - \left(\sum_{i \in \mathcal{K}} H(W_i)_{\psi'} - H(W_{\mathcal{K}})_{\psi'} \right) - f_1(\epsilon, n) \\
&\geq \left[\sum_{i \in \mathcal{K}} H(A_i) + H(R^n) - H(R^n A_{\mathcal{K}}) \right] \\
&\quad - 2E_{\text{sq}}(A_{k_1}; \dots; A_{k_l})_{\psi} - f_1(\epsilon, n) - f'_2(\epsilon, n).
\end{aligned}$$

We can simplify the expression further by using the fact that $|\psi\rangle = |\varphi\rangle^{\otimes n}$ to obtain

$$\begin{aligned}
\sum_{k \in \mathcal{K}} Q_k &\geq \frac{1}{2} \left[\sum_{k \in \mathcal{K}} H(A_k) + H(R) - H(RA_{\mathcal{K}}) \right] \\
&\quad - E_{\text{sq}}(A_{k_1}; A_{k_2}; \dots; A_{k_l})_{\varphi} - \frac{f_1(\epsilon, n)}{2n} - \frac{f'_2(\epsilon, n)}{2n}
\end{aligned}$$

where the we used explicitly the additivity of the entropy for tensor product states and the subadditivity of squashed entanglement demonstrated in Proposition 4.7.

Theorem 5.3 follows from the above since $\epsilon > 0$ was arbitrary and the sum $(f_1(\epsilon, n) + f'_2(\epsilon, n))/n \rightarrow 0$ as $\epsilon \rightarrow 0$. \square

5.4 Discussion

The multiparty fully quantum Slepian-Wolf protocol is an optimal solution to the distributed compression problem for separable states, i.e. states of the form

$$\varphi^{X_1 \cdots X_m} = \sum_i p_i \varphi_i^{X_1} \otimes \varphi_i^{X_2} \otimes \cdots \otimes \varphi_i^{X_m},$$

because $E_{\text{sq}} = 0$ for such states. For general states, we have provided an outer bound on the set of achievable rates based on the multiparty squashed entanglement. In this section, we outline some other aspects of the multiparty FQSW protocol and its relation to other protocols.

First, we note that there is an alternative, more compact way of writing the rate sum inequalities of Theorem 5.2 and Theorem 5.3. Consider the inequalities of the inner bound (5.2) reproduced below:

$$\sum_{k \in \mathcal{K}} Q_k \geq \frac{1}{2} \left[\sum_{k \in \mathcal{K}} H(A_k) + H(R) - H(RA_{\mathcal{K}}) \right], \quad \forall \mathcal{K} \subseteq \{1, \dots, m\}. \quad (5.44)$$

The term on the right hand side can be expressed as a multiparty information

$$\sum_{k \in \mathcal{K}} Q_k \geq \frac{1}{2} I(A_{\mathcal{K}}; R), \quad \forall \mathcal{K} \subseteq \{1, \dots, m\}, \quad (5.45)$$

where $I(A_{\mathcal{K}}; R)$ is the multiparty information of all the members of \mathcal{K} and R . The multiparty information function is naturally suited to the multiparty distributed compression problem.

When only two parties are involved ($m = 2$), the inequalities in (5.44) reduce to the two-party bounds on distributed compression presented in [5]:

$$\begin{aligned} Q_1 &\geq \frac{1}{2} I(A_1; R), \\ Q_2 &\geq \frac{1}{2} I(A_2; R), \\ Q_1 + Q_2 &\geq \frac{1}{2} [H(A_1) + H(A_2) + H(A_1 A_2)]. \end{aligned} \quad (5.46)$$

However, we now understand the mystery behind the expression that looks like the mutual information with a reversed sign: it is simply the form $\frac{1}{2}I(A_1; A_2; R)$, where $H(R) = H(A_1A_2)$ and $H(A_1A_2R) = 0$. The outer bound inequalities (5.3) similarly reduce to the corresponding expressions in the FQSW paper [5] with the multiparty squashed entanglement being replaced by the original two-party squashed entanglement of [17].

Another observation concerns the classical communication cost of the protocol. If we move away from the “fully quantum” regime and allow classical communication between the senders and the receiver we can achieve better rates. We do this by recycling the entanglement generated by the FQSW protocol. For two parties, the combination of multiparty FQSW of equation (5.46) with teleportation reproduces the state merging results of equation (3.21)

$$\begin{aligned} R_A &> H(A|B)_\rho, \\ R_B &> H(B|A)_\rho, \\ R_A + R_B &> H(AB)_\rho. \end{aligned} \tag{5.47}$$

Finally we note that the multiparty FQSW protocol can be operated backwards in time to produce an optimal reverse Shannon theorem for the quantum broadcast channel [57].

CHAPTER 6

Possible applications to the black hole information paradox

There are very few physical systems that require both the application of the principles of general relativity and of quantum mechanics in order to understand them. Black holes fall into this category. Classically, a black hole is a region of space where the gravity is so strong that nothing can escape its pull – not even light. However, according to a certain semi-classical calculation performed by Hawking [73], black holes emit thermal radiation at a very slow rate. Thus, while it may take a very long time, all the mass/energy that fell into the black hole will eventually be released back into the universe and the black hole will evaporate.

This scenario poses a serious problem known as the black hole information paradox. Consider a universe originally in the pure state $|\text{Universe}\rangle$ which collapses onto itself to form a black hole. After a very long time, the black hole evaporates completely to leave behind a universe filled with thermal radiation, which corresponds to the maximally mixed state. Herein lies the paradox: an initially pure state has evolved to a mixed state — something which violates the laws of unitary evolution so central to all of quantum theory.

Does gravity lead to non-unitary evolution or is general relativity incomplete? Over the last 30 years, many preeminent physicists have had something to say about this question and yet this paradox still defies explanation [74, 75, 76]. What is worse is that the more we think about the information paradox the more we realize that it is not an “unwarranted extrapolation from an untrustworthy approximation” [74] but rather a true paradox of physics that

cannot be explained yet. True paradoxes of this kind are indicators that the scientific theories we use do not provide a complete description of reality.

The black hole information paradox is yet to be explained in a satisfactory manner by modern physics and perhaps will not be until a theory of quantum gravity is developed. Recently, however, interesting contributions to the black hole information problem have been made by people from within the quantum information community [77, 78, 79, 80, 81]. In the last chapter of this thesis, we present a curious and counter-intuitive result about the nature of purifications and then use this observation to make a speculative comment about black holes with highly mixing internal dynamics.

6.1 Polygamy of purification

In a closing remark of the original FQSW paper [5], the authors make a very interesting observation about the nature of quantum purifications which we will refer to as *polygamy of purification*. Consider three parties — Alice, Bob and Ron who share the quantum state

$$|\psi\rangle^{A^n B^n R^n} = \left(|\Phi\rangle^{A_B B} \otimes |\Phi\rangle^{A_R R} \right)^{\otimes n}, \quad (6.1)$$

where $|\Phi\rangle$ denotes the maximally entangled state $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. In other words, Alice shares n entangled states with Ron and another n maximally entangled states with Bob. The entanglement structure is illustrated in Figure 6–1 a).

Now, we tell Alice to perform the standard FQSW task, that is, to transfer her R entanglement to Bob. Suppose that Alice performs the standard FQSW

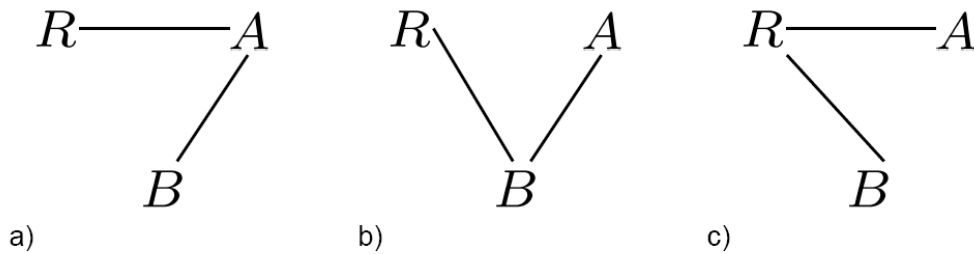


Figure 6–1: Transfer of quantum correlations between three parties: (a) The original AR and AB entanglement. (b) The effect of Alice sending the system A_1 to Bob. She is completely decoupled from the R system. (c) Alternatively, Alice can send the *same* A_1 system to Ron and completely decouple from Bob!

protocol in order to accomplish the entanglement transfer¹. She applies a random unitary to the system A^n and then sets aside a subsystem A_1 of dimension d_{A_1} where

$$\log d_{A_1} \geq \frac{1}{2}I(A; R)_\phi = n \quad [\text{qubits}] \quad (6.2)$$

as required by equation (3.25) for the FQSW protocol. Sending the system A_1 to Bob will successfully decouple Alice from Ron and lead to the entanglement configuration illustrated in Figure 6–1 b).

Note, however, that the encoding operation was not specifically targeting Bob. Indeed, if the same A_1 system is sent to Ron instead, we would transfer the Bob entanglement to him and obtain the configuration of Figure 6–1 c). The polygamy of purification, therefore, is the observation that it is possible for a single quantum system A_1 to contain the purification of more than one other system!

¹ Since in our setup the R -entangled part of her system is clearly identifiable, another approach for Alice could be to simply take the n Ron-entangled qubits and send them to Bob.

6.2 Random internal dynamics for black holes

Recently, the results of the FQSW protocol were connected to the black hole information paradox [79]. The question studied is not about the evolution of the universe as whole but something more specific. If we drop half of pure state $|\varphi\rangle^{AB_1}$ into a black hole, denoted B_2 , how long will it take for the its purification to come out?

Under the assumption that the internal dynamics of the black hole correspond to a random unitary operation, a situation which was considered previously in [76], we can give an answer to this question since it corresponds to an FQSW-type of problem except for the Schumacher compression step. We model the internal black hole dynamics as a random unitary U_B which takes the system $B = B_1B_2$ to an isomorphic system $B'R$, where R is released as radiation and B' is what remains of the black hole. The rest of the universe is denoted U and no assumptions are made about its size. The situation is illustrated in Figure 6–2.

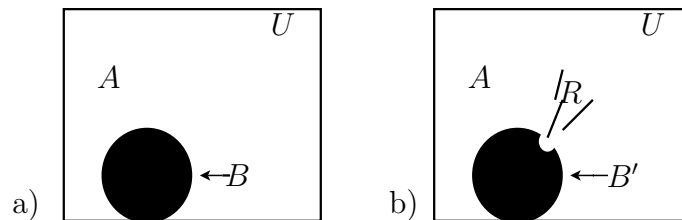


Figure 6–2: **a)** Black hole before the radiative process has taken place. The purification of the A system, B_1 , is somewhere inside the black hole. The system U denotes the rest of the universe, i.e. everything that is not A or B . **b)** After the black hole emits the radiation chunk R the remainder of the black hole is labeled B' .

Inspired by the FQSW results, we can say that if the dimension of the radiated system satisfies

$$\log d_R \geq \frac{1}{2}I(A; B) = \frac{1}{2}I(A; B_1) = H(A) \quad (6.3)$$

then, with high probability, it will contain the purification of the A system. This is because we can think of the black hole as an active entity mixing its internal degrees of freedom.

In the current setup, we do not have the luxury of working in the i.i.d. regime so the statements we make are nothing more than inspired hand waving arguments. Nevertheless, our calculation leads us to speculate that the purification information of a specific system will come out fairly fast and independently of the size of the black hole. In fact, since the system we labeled A was arbitrary, the purification of all subsystems of the universe with the same dimension comes out with the radiation R ! This is not so surprising since we already know about the polygamy of purification. Nevertheless, even if the purification of any particular system of interest comes out quickly, we still have to wait until all of the black hole evaporates to recover the the purification of the whole universe, so the original black hole paradox remains.

It is not clear what we mean when we say that the black hole has “internal dynamics”. To assume that something interesting happens at the horizon is OK perhaps, but aren’t black holes supposed to trap systems forever?

6.3 Lost subsystem problem

Consider now a similar situation to the above but this time the black hole consists of two systems B_2L , where the L system is “lost”; nothing ever leaves L . Half of a pure state $|\varphi\rangle^{AB_1}$ is dropped into the black hole which is assumed to have random unitary dynamics on the space $B = B_1B_2$ from which a system R is emitted. Once more we label B' the remainder of the black hole as illustrated in Figure 6–3.

We would like to know how big the R system has to be in order for the purification of A to come out. This time, there are two active “participants”: B and L , so the multiparty FQSW results have to be considered. Thus, in

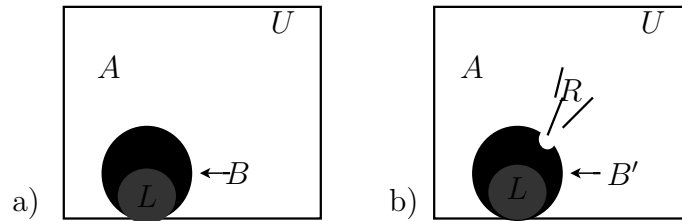


Figure 6–3: **a)** The lost subsystem L is part of the black hole BL . The system U denotes the rest of the universe. **b)** The black hole has released radiation R from the B subsystem. The remainder of the black hole is $B'L$.

order for the purification of A to come out the dimension of the radiated systems have to satisfy

$$\begin{aligned}
 \log d_R &\geq \frac{1}{2}I(B; A) = \frac{1}{2}I(B_1; A) = H(A), \\
 \log d_{R_L} &\geq \frac{1}{2}I(L; A) = 0, \\
 \log d_R + \log d_{R_L} &\geq \frac{1}{2}I(L; B; A) = H(A) + \frac{1}{2}I(B_2; L).
 \end{aligned} \tag{6.4}$$

where d_{R_L} is the dimension of the system released by the lost system.

At first sight, all seems to be in order since the requirement $\log d_{R_L} \geq 0$ is satisfied. The inequality for the sum of the rates, however, adds an extra requirement for d_R . To see the purification of A come out we will have to wait until

$$\log d_R > \max\{H(A), H(A) + \frac{1}{2}I(B_2; L)\}. \tag{6.5}$$

Thus, if there are any significant correlations between the B_2 and L parts of the black hole the information will *not* come out quickly. This result is very interesting because the purification of A will be slow to come out even though it is held in the B part of the black hole and hasn't completely fallen into the L system.

CHAPTER 7

Conclusion

This thesis has been an expedition into the field of quantum information science with many twists and turns. We began by introducing the fundamental principles of classical information theory and their extensions to the quantum realm. Armed with the basics, we were ready to approach some of last decade's important results in quantum information theory with the aim of getting readers from outside the field up to speed.

We then attacked the multiparty distributed compression problem with the most powerful weapon available in our arsenal: the fully quantum Slepian-Wolf protocol. The construction of the multiparty distributed compression protocol is conceptually simple. It consists of sequential applications of the two-party FQSW protocol with careful accounting of the information theoretic quantities at each step. However, in order to achieve rigorous proofs of the bounds on the multiparty rate region, we had to wage a heavy battle in difficult but interesting terrain.

To achieve a rigorous proof of Theorem 5.2, the inner bound on the rate region, we had to dig into the geometry of convex polyhedra in m -dimensional space. The proof we obtained uses a sufficient level of mathematical abstraction so as to apply to other problems in information theory involving multiparty rate regions proved in terms of achievable points but expressed instead in terms of facet inequalities. Indeed, our proof is valid for all supermodular rate regions, that is, all rate region specified by a set of inequalities

$$\sum_{k \in \mathcal{K}} R_k \geq C_{\mathcal{K}}, \quad \forall \mathcal{K} \subseteq \{1, \dots, m\} \quad (7.1)$$

for which the constants C_K satisfy the supermodular condition $C_{K \cup \mathcal{L}} + C_{K \cap \mathcal{L}} \geq C_K + C_{\mathcal{L}}$. In particular, the rate regions for the classical multiparty Slepian-Wolf problem [33, 34] and the multiparty state merging protocol [43] fall into this category because of strong subadditivity.

Also, in order to prove Theorem 5.3, the outer bound on the rate region, it was necessary to formulate a definition of the multiparty information and from it derive a multiparty generalization of the squashed entanglement. In the chapter dedicated to the multiparty squashed entanglement, we showed that it is a continuous, convex and subadditive measure of entanglement — all desirable but rare properties in the multiparty case.

Some open problems remain which could form fruitful directions for future investigations. The additivity of the multiparty squashed entanglement is an important conjecture that was recently proved in an updated version of [16], which now includes W. Song in the author list. As for the distributed compression problem, we have fully solved the problem only for separable states. Perhaps a different correction term exists for the outer bound? If we find states for which we can calculate E_{sq} analytically or numerically we could use them to further probe the shape of the outer bound. Of course, the black hole information paradox remains an open problem since it hasn't been solved by our toy-model observations.

And so, we add the new *weapon of mass decoupling* to the ever growing collection of quantum information theory protocols derived from the nearly-universal building block of two-party FQSW. At the time of writing of this thesis, this collection contains entanglement distillation, channel simulation, communication over quantum broadcast channels, and many others. In fact, even the more general state redistribution [55] result can be obtained from the FQSW protocol [82].

References

- [1] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Jl. Res. Develop.*, 5:183, 1961.
- [2] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53(4):2046–2052, 1996. arXiv:quant-ph/9511030.
- [3] I. Devetak, A. W. Harrow, and A. Winter. A family of quantum protocols. *Phys. Rev. Lett.*, 93:230504, 2004. arXiv:quant-ph/0308044.
- [4] I. Devetak, A. W. Harrow, and A. Winter. A resource framework for quantum Shannon theory. 2005. arXiv:quant-ph/0512015.
- [5] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information’s family tree. 2006. arXiv:quant-ph/0606225.
- [6] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44(1):269–273, 1998. arXiv:quant-ph/9611023.
- [7] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997. doi:10.1103/PhysRevA.56.131.
- [8] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.*, 83:3081, 1999. arXiv:quant-ph/9904023.
- [9] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44, 2005. arXiv:quant-ph/0304127.
- [10] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824, 1996. arXiv:quant-ph/9604024.
- [11] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. London A*, 461:207–235, 2005. arXiv:quant-ph/0306078.

- [12] M. B. Plenio and S. Virmani. An introduction to entanglement measures. *Quant. Inf. Comp.*, 7:1, 2007. arXiv:quant-ph/0504163.
- [13] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4):471–480, 1973.
- [14] C. Ahn, A. Doherty, P. Hayden, and A. Winter. On the distributed compression of quantum information. *IEEE Trans. Inf. Theory*, 52:4349, 2006. arXiv:quant-ph/0403042.
- [15] M. Horodecki, J. Oppenheim, and A. Winter. Quantum information can be negative. *Nature*, 436:673, 2005. doi:10.1038/nature03909.
- [16] D. Yang, K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, and W. Song. Squashed entanglement for multipartite states and entanglement measures based on the mixed convex roof. 2007. arXiv:0704.2236.
- [17] M. Christandl and A. Winter. Squashed entanglement - an additive entanglement measure. *J. Math. Phys.*, 45:829, 2004. arXiv:quant-ph/0308088.
- [18] P. M. Hayden, M. Horodecki, and B. M. Terhal. The asymptotic entanglement cost of preparing a quantum state. *J. Phys. A: Math. Gen.*, 34:6891–6898, 2001. doi:10.1088/0305-4470/34/35/314.
- [19] E. M. Rains. A rigorous treatment of distillable entanglement. *Phys. Rev. A*, 60:173, 1999. arXiv:quant-ph/9809078.
- [20] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619, 1998. arXiv:quant-ph/9707035.
- [21] N. Linden, S. Popescu, B. Schumacher, and M. Westmoreland. Reversibility of local transformations of multiparticle entanglement. *Quant. Inf. Proc.*, 4(3):241–250, 2005. arXiv:quant-ph/9912039.
- [22] W. Dur, J. I. Cirac, and R. Tarrach. Separability and distillability of multiparticle quantum systems. *Phys. Rev. Lett.*, 83:3562, 1999. arXiv:quant-ph/9903018.
- [23] V. Coffman, J. Kundu, and W. K. Wootters. Distributed entanglement. *Phys. Rev. A*, 61:052306, 2000. arXiv:quant-ph/9907047.
- [24] C. H. Bennett, S. Popescu, D. Rohrlich, J. A. Smolin, and A. V. Thapliyal. Exact and asymptotic measures of multipartite pure-state entanglement. *Phys. Rev. A*, 63(1):012307, Dec 2000. arXiv:quant-ph/9908073.
- [25] D. Avis, P. Hayden, and I. Savov. Multiparty distributed compression and squashed entanglement. 2007. arXiv:0707.2792.

- [26] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- [27] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.
- [28] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 623–656, 1948.
- [29] P. Mirowski. *Machine dreams, economics becomes a cyborg science*. Cambridge University Press, 2001.
- [30] H. B. Callen. *Thermodynamics and an introduction to thermostatistics*. John Wiley & Sons, 1985.
- [31] M. C. Mackey. *Time’s Arrow: The origins of thermodynamic behavior*. Springer-Verlag, 1992.
- [32] I. Csiszár and J. Körner. *Information Theory: Coding theorems for discrete memoryless systems*. Akadémiai Kiadó, Budapest, 1981.
- [33] J. Wolf. Data reduction for multiple correlated sources. *Proc. 5th Colloquium Microwave Comm.*, pages 287–295, 1974.
- [34] T. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Trans. Inf. Theory*, 21(2):226–228, 1975.
- [35] J.J. Sakurai. *Modern quantum mechanics*. Addison-Wesley, 1994.
- [36] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.*, 14:1938–1941, 1973.
- [37] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. 2007. arXiv:quant-ph/0702225.
- [38] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.
- [39] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.
- [40] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Theory*, 45:1216, 1999. doi:10.1109/18.761271.
- [41] B. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614–2628, 1996. arXiv:quant-ph/9604023.

- [42] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51:2738–2747, 1995. doi:10.1103/PhysRevA.51.2738.
- [43] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. arXiv:quant-ph/0512247, 2005.
- [44] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. Lond. A*, 461:207235, 2005. arXiv:quant-ph/0306078.
- [45] M. Horodecki, P. Horodecki, R. Horodecki, D. W. Leung, and B. M. Terhal. Classical capacity of a noiseless quantum channel assisted by noisy entanglement. *Quant. Inf. Comp.*, 1(3):7078, 2001. arXiv:quant-ph/0106080.
- [46] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55(3):16131622, 1997.
- [47] P. W. Shor. The quantum channel capacity and coherent information. *MSRI workshop on quantum computation*, 2002.
- [48] I. Devetak. Triangle of dualities between quantum communication protocols. *Phys. Rev. Lett.*, 97(14):140503, 2006. arXiv:quant-ph/0505138.
- [49] N. J. Cerf and C. Adami. Negative entropy and information in quantum mechanics. *Phys. Rev. Lett.*, 79(26):5194–5197, 1997. arXiv:quant-ph/9512022.
- [50] N.J. Cerf and C. Adami. Information theory of quantum entanglement and measurement. *Physica D*, 120:62–81, 1998.
- [51] N. J. Cerf and C. Adami. Entropic Bell inequalities. *Phys. Rev. A*, 55(5):3371–3374, 1997.
- [52] D. P. DiVincenzo, C. A. Fuchs, H. Mabuchi, J. A. Smolin, A. Thapliyal, and A. Uhlmann. Entanglement of assistance. *Lect. Notes Comp, Sci.*, 1509:247–257, 1999. arXiv:quant-ph/9803033.
- [53] F. Dupuis and P. Hayden. A father protocol for quantum broadcast channels. 2006. arXiv:quant-ph/0612155.
- [54] I. Devetak and J. Yard. The operational meaning of quantum conditional information. 2006. arXiv:quant-ph/0612050.
- [55] J. Yard and I. Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. 2007. arXiv:0706.2907.
- [56] D. Leung, J. Oppenheim, and A. Winter. Quantum network communication – the butterfly and beyond, 2006. arXiv:quant-ph/0608223.

- [57] P. Hayden and F. Dupuis. An optimal reverse shannon theorem for quantum broadcast channels. In preparation, 2007.
- [58] G. Lindblad. Entropy, information and quantum measurements. *Commun. Math. Phys.*, 33:305–322, December 1973.
- [59] R. Horodecki. Informationally coherent quantum systems. *Phys. Lett. A*, 187:145–150, April 1994. doi:10.1016/0375-9601(94)90052-3.
- [60] B. Groisman, S. Popescu, and A. Winter. Quantum, classical, and total amount of correlations in a quantum state. *Phys. Rev. A*, 72(3):032317, 2005.
- [61] R. R. Tucci. Quantum entanglement and conditional information transmission. 1999. arXiv:quant-ph/9909041.
- [62] R. R. Tucci. Entanglement of distillation and conditional mutual information. 2002. arXiv:quant-ph/0202144.
- [63] U. Maurer and S. Wolf. Unconditionally secure key agreement and the intrinsic conditional information. *IEEE Trans. on Inf. Theory*, 45(2):499–514, 1999.
- [64] M. Christandl. *The structure of bipartite quantum states. Insights from group theory and cryptography*. PhD thesis, Selwyn College, University of Cambridge, 2006.
- [65] G. Vidal. Entanglement monotones. *J. Mod. Opt.*, 47:355, 2000. arXiv:quant-ph/9807077.
- [66] E. Davies and J. Lewis. An operational approach to quantum probability. *Commun. Math. Phys.*, 17:239–260, 1970.
- [67] R. Alicki and M. Fannes. Continuity of quantum mutual information. 2003. arXiv:quant-ph/0312081.
- [68] A. Uhlmann. The ‘transition probability’ in the state space of a $*$ -algebra. *Rep. Math. Phys.*, 9:273, 1976.
- [69] G. M. Ziegler. *Lectures on polytopes*. Springer-Verlag, New York, 1995.
- [70] J. Edmonds. Submodular functions, matroids, and certain polyhedra. *Proc. Calgary Int. Conf. Combinatorial Structures and Algorithms*, pages 69–87, June 1969. (Reprinted in *LNCS* 2570:11–26, 2003).
- [71] P. Hayden and A. Winter. Achievable rates for multiparty distributed compression. Unpublished, 2006.

- [72] D. Tse and S. Hanley. Multiaccess fading channels: Polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Trans. Inf. Theory*, 44(7):2796–2815, 1998.
- [73] S. W. Hawking. Particle creation by black holes. *Comm. Math. Phys.*, 43:199–220, 1975.
- [74] J. Preskill. Do black holes destroy information? arXiv:hep-th/9209058.
- [75] J. Traschen. An introduction to black hole evaporation. arXiv:gr-qc/0010055.
- [76] D. N. Page. Hawking radiation and black hole thermodynamics. arXiv:hep-th/0409024.
- [77] C. Adami and G.L. Ver Steeg. Black holes conserve information in curved-space quantum field theory. 2004. arXiv:gr-qc/0407090.
- [78] C. Adami and G.L. Ver Steeg. Black holes are almost optimal quantum cloners. 2006. arXiv:quant-ph/0601065.
- [79] P. Hayden and J. Preskill. Black holes as mirrors: quantum information in random subsystems. 2007. In preparation.
- [80] J. Smolin and J. Oppenheim. Information locking in black holes. *Phys. Rev. Lett.*, 96:081302, 2006. arXiv:hep-th/0507287.
- [81] R. Buniy and S. Hsu. Entanglement entropy, black holes and holography. *Phys. Lett. B*, 644:72, 2007.
- [82] J. Oppenheim. Redistributing quantum information from fully quantum Slepian-Wolf. Private communication, 2007.